

**Development and Evaluation of Biomarkers in
Huntington's Disease:
Furthering our Understanding of the Disease and
Preparing for Clinical Trials**

Thesis submitted for the degree of Doctor of Philosophy

Elin M Rees

Institute of Neurology

University College London

2014

Declaration

I, Elin Rees, confirm that the work presented in this thesis is my own work. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Elin Rees

Abstract

Huntington's Disease (HD) is a devastating hereditary neurodegenerative disease for which there are currently only symptomatic treatments. Several potentially curative pharmaceutical and genetic therapies are however in varying stages of development and therefore an increasing number of large-scale clinical trials of disease-modifying therapies are imminent. There is consequently a need for biomarkers which are sensitive to beneficial attenuation of disease-related changes.

Functional, neuroimaging and biochemical biomarkers have been developed in HD (Andre et al. 2014;Weir et al. 2011). Neuroimaging biomarkers are strong candidates based on their clear relevance to the neuropathology of disease, proven precision and superior sensitivity compared with some standard functional measures (Tabrizi et al. 2011;Tabrizi et al. 2012). Their use in early-stage clinical trials, as surrogate end-points providing initial evidence of biological effect, is becoming increasingly common. Comparison of biomarkers in HD will help to clarify which measures, over varying time intervals, are most sensitive to disease progression. Additionally, the identification of robust fully-automated methods, comparable to manual and semi-automated gold-standards, would facilitate large-scale volumetric analysis. These methods however require validation in observational studies of neurodegenerative disease before they can be applied to sensitive clinical trial data.

This thesis will develop and evaluate biomarkers for use in HD; both furthering our understanding of the disease and in preparation for use as end-points in clinical trials. A direct comparison of the sensitivity of diffusion and volumetric imaging biomarkers to HD-related change will be reported for the first time. Several exploratory imaging investigations are also described which enhance current knowledge of the relationship between neuroimaging metrics, brain functioning and behaviour, additionally strengthening the argument for the clinical relevance of neuroimaging measures as surrogate end-points in HD. The thesis will conclude with a comprehensive biomarker evaluation in early-stage HD, along with suggested strategies for selection of primary and secondary trial end-points based on effect sizes and corresponding sample size requirements.

Table of Contents

Declaration	1
Abstract	2
Table of Contents	3
Table of Figures	9
Table of Tables.....	13
List of Abbreviations	15
Aims of this Thesis	17
1 Background: Huntington’s Disease (HD)	18
1.1 Genetics.....	18
1.2 Neuropathology.....	19
1.3 Clinical Presentation	19
1.4 Treatment Scope	20
1.5 Research Methods.....	21
1.5.1 Clinical Rating Scales.....	21
1.5.2 Biomarkers.....	23
1.5.3 Research Cohorts.....	23
1.5.4 Neuroimaging Study Set-Up	24
2 Neuroimaging in HD: A Literature Review.....	25
2.1 Structural MRI.....	25
2.1.1 Basal Ganglia.....	25
2.1.2 Whole-Brain.....	27
2.1.3 Ventricles	29
2.1.4 Grey and White Matter	29
2.1.5 Correlations with Clinical Measures	30
2.1.6 Summary: Structural MRI in HD.....	30
2.2 Diffusion MRI	32
2.2.1 Reported Findings.....	33
2.2.2 Summary: Diffusion MRI in HD	34
2.3 Functional MRI (fMRI)	34
2.4 PET	35
2.5 Magnetic Resonance Spectroscopy (MRS)	35
2.6 Neuroimaging Biomarkers in Clinical Trials	35
2.7 Neuroimaging in HD: Literature Summary	35
3 Thesis Methods: The PADDINGTON Study – Work Package 2	38

3.1	Cohort.....	38
3.2	Ethical Approval.....	39
3.3	Assessments	39
3.4	Image Acquisition	40
4	Thesis Methods: Image Analysis.....	42
4.1	Image Registration.....	42
4.1.1	Target of Registration	42
4.1.2	Transformation Models	43
4.1.3	Resampling and Interpolation	45
4.1.4	Goodness-of-Fit	46
4.1.5	Exit Criteria	46
4.2	Structural MR Image Analysis.....	46
4.2.1	Pre-Processing	46
4.2.2	Manual Delineation	48
4.2.3	Automated Analysis.....	51
4.2.4	FreeSurfer	51
4.2.5	FIRST	52
4.2.6	BRAINS	52
4.2.7	STEPS	52
4.2.8	Fluid Registration.....	53
4.2.9	Statistical Parametric Mapping (SPM).....	53
4.2.10	GM and WM Atrophy	55
4.2.11	The Boundary Shift Integral (BSI).....	55
4.3	Diffusion MR Image Analysis	57
4.3.1	Pre-Processing	57
4.3.2	ROI Analysis	59
4.3.3	Longitudinal ROI Analysis	60
5	Thesis Methods: Statistical Analyses.....	64
5.1	Regression	64
5.2	Covariates	64
5.3	Effect and Sample Size Calculations	64
5.4	Correction for Multiple Comparisons.....	65
6	PART 1. Development, Evaluation and Clinical Application of Tools Sensitive to HD-Related Pathology	
	67	
7	Development, Evaluation and Application of Tools Sensitive to Striatal Atrophy in HD	68

7.1	Background.....	68
7.2	Development of a Manual Delineation Protocol for the Putamen in HD	69
7.2.1	Selection of Intensity Thresholds	69
7.2.2	Definition of Inferior Cut-off.....	69
7.2.3	Reproducibility.....	71
7.2.4	Inter- and Intra-Scanner Type Tests	72
7.2.5	Protocol Development: A Summary	75
7.3	A Method Comparison	76
7.3.1	Aim.....	76
7.3.2	Methods.....	76
7.3.3	Results	78
7.3.4	Discussion	93
8	Volumetric Analysis of the Cerebellum	96
8.1	Background.....	96
8.2	Cohort: TRACK-HD	96
8.3	Protocol Development.....	96
8.3.1	Voxel-Intensity Thresholds	96
8.3.2	Boundary Definition.....	99
8.3.3	Statistical Comparison of Thresholds	99
8.3.4	Geometric Distortion.....	103
8.4	Cerebellar Volumetry in HD: A Method Comparison	104
8.4.1	Background.....	104
8.4.2	Cerebellar KN-BSI Development.....	105
8.4.3	Methods.....	106
8.4.4	Results	108
8.4.5	Discussion	109
9	Evaluation and Application of FreeSurfer Cortical Thickness Analysis as a Biomarker in HD	111
9.1	Introduction.....	111
9.2	Image Analysis	111
9.3	Statistical Analysis	112
9.4	Methodological Investigations of FreeSurfer Cortical Thickness Software.....	113
9.4.1	A Comparison of Change Metrics	113
9.4.2	The Effect of Increasing the Smoothing Kernel	118
9.4.3	The Effect of Adjusting for Multiple Comparisons	122
9.4.4	The Differences between FreeSurfer Versions (5.1.0 vs 5.3.0)	124

9.5	Cortical Thinning as a Marker of Disease Progression in Early HD.....	128
9.5.1	Background.....	128
9.5.2	Methods.....	130
9.5.3	Results	131
9.5.4	Discussion	136
10	Development and Administration of a Novel Cognitive Test: Multi-Modal Emotion Recognition in HD	138
10.1	Background.....	138
10.2	Task Design	140
10.2.1	Stimuli.....	140
10.2.2	Paradigm.....	141
10.2.3	Control Tasks	141
10.3	Application to an HD Cohort.....	141
10.4	Statistical Analysis	142
10.5	Results	142
10.5.1	Facial Recognition.....	142
10.5.2	Emotion Recognition	143
10.6	Discussion	147
11	PART 1: A Summary	150
12	PART 2. Exploratory Investigations of Clinical, Cognitive and Neuroimaging Associations in HD	152
13	Cerebellar Abnormalities in HD: A Role in Motor and Psychiatric Impairment?	153
13.1	Introduction.....	153
13.2	Methods	154
13.2.1	Cohort.....	154
13.2.2	Image Acquisition	154
13.2.3	Image Pre-Processing	155
13.2.4	Image Analysis	155
13.2.5	Voxel-Based Morphometry (VBM)	156
13.2.6	Motor and Psychiatric Assessments.....	157
13.2.7	Statistical Analysis	157
13.3	Results	157
13.4	Discussion	162
14	Emotion Recognition in HD: An Exploratory Imaging Investigation	165
14.1	Background.....	165
14.2	Methods	166

14.2.1	Image Acquisition	166
14.2.2	Image Analysis	166
14.2.3	Statistical Analysis	166
14.3	Results	167
14.3.1	Macro-Structural Associations	167
14.3.2	Micro-Structural Associations	168
14.4	Discussion	176
15	A Silent Contribution of the Visual Cortex to Cognitive Task Performance: An Investigation in HD 179	
15.1	Introduction.....	179
15.2	Methods	180
15.2.1	Cohort.....	180
15.2.2	Clinical Assessments	181
15.2.3	Image Acquisition	182
15.2.4	Image Analysis	182
15.2.5	Statistical Analysis	183
15.3	Results	183
15.4	Discussion	188
16	PART 2: A Summary	190
17	PART 3. Evaluation of Neuroimaging Measures as Biomarkers in HD	191
18	Evaluation of Multi-Modal, Multi-Site Neuroimaging Measures as Biomarkers in HD: Results from the PADDINGTON study	196
18.1	Background.....	196
18.1.1	Micro- versus Macro-Structural ROI-Based Biomarkers.....	196
18.1.2	Cortical Thickness	196
18.1.3	Short-Interval VBM	197
18.2	Methods	197
18.2.1	Cohort.....	197
18.2.2	Image Acquisition and Analysis	198
18.2.3	Cognitive and Clinical Assessments	200
18.2.4	Statistical Analysis	200
18.3	Results	200
18.3.1	Short-Interval VBM	207
18.4	Discussion	208
18.5	Guidelines for Clinical Trial Biomarker Selection in Early-Stage HD.....	210
19	PART 3: A Summary	213

20	Conclusions.....	214
21	Publications	216
22	Acknowledgements	219
23	Appendix.....	220
23.1	UHDRS: TMS Scoring System	220
23.2	UHDRS: TFC Scoring System	224
23.3	TRACK-HD Acquisition Parameters.....	224
23.4	Manual Putamen Delineation: Standard Operational Procedure (SOP)	224
23.5	Manual Cerebellum Delineation: SOPs.....	228
23.5.1	Protocol 1	228
23.5.2	Protocol 2	232

Table of Figures

Figure 1-1. A model of the progression of HD neuropathology over a patient's lifespan from preHD through diagnosis (represented as a dotted line) to manifest disease. Adapted from Ross & Tabrizi (Ross and Tabrizi 2011).....	19
Figure 1-2. A model of the progression of HD symptoms over a patient's lifespan. Adapted from Ross & Tabrizi (Ross & Tabrizi 2011).	19
Figure 4-1. (Left) An example of a good quality VCM from one participant following fluid registration. The GM and WM can be seen to be contracting (green-blue) and the ventricular and CSF space expanding (yellow-red). (Right) A poor quality VCM showing significant distortion particularly of the cerebellar and brainstem regions.....	45
Figure 4-2. Voxel intensity fluctuations before (left) and after (right) N3 bias correction.	47
Figure 4-3. a) An example of motion artefacts, in this case ringing and blurring; b) one participant's scans at baseline and follow-up, with inconsistent head positioning in the FOV.....	48
Figure 4-4. a) A screen shot of a whole-brain segmentation in MIDAS software. b) Examples of regional segmentations of (top left to right) whole-brain, lateral ventricles, putamen, (bottom left to right) caudate, corpus callosum and total intracranial volume.	49
Figure 4-5. a) FreeSurfer volumetric segmentations and cortical delineation displayed with Tkmedit, b) FIRST and c) BRAINS3 subcortical segmentations and d) SPM GM and WM segmentations, all displayed in FSLView and e) a STEPS caudate segmentation displayed in MIDAS software.....	51
Figure 4-6. Examples of frontal distortions in raw diffusion data (top) and hyper-intensities in regions proximal to the nasal sinuses (below).	58
Figure 4-7. An overview of the cross-sectional PADDINGTON diffusion analysis pipeline.	60
Figure 4-8. An overview of the longitudinal PADDINGTON diffusion analysis. 1) The baseline T1 was affine registered to the follow-up image and vice versa and then a symmetric 'average' transformation was calculated between the two time-points. With the binary regions from baseline and follow-up now in common space the two were multiplied together, in order to 'zero' (i.e. remove) any voxels not present at both time-points, only retaining those common to both baseline and follow-up; 2) the same process was applied to the FA maps; 3) the T1 half-way images and ROI were then warped into FA half-way space; 4) using the inverse of the B0 native-to-half-way-space transformation, the T1 images were moved into native diffusion space for each of the baseline and follow-up and regional means for FA, MD, RD and AD were generated within these regions.....	62
Figure 5-1. Sample size and statistical power calculations for different effect sizes (taken from Tabrizi et al. (Tabrizi et al. 2012))......	65
Figure 7-1. Putamen segmentation protocol examples of the initially thresholding stage: a) Some lighter GM voxels may be excluded by the intensity thresholds and therefore several seeds may be needed to roughly fill the structure on each slice; b) Darker WM voxels may also be included and therefore manual editing is often required down the lateral border (example in green), disconnecting the putamen from the WM and claustrum.....	69
Figure 7-2. Putamen protocol examples: the most inferior slice in which the putamen should be segmented (left) – here the putamen is still clearly separated by WM from the caudate (red arrows). Once this separation is no longer clear (e.g. images on the right) stop the segmentation and do not segment this slice.	70

Figure 7-3. Putamen protocol example: if the tail and head of the putamen are connected on that slice segment the whole structure (left). When this becomes split by WM deseed the tail (right). Dark spots within the structure (e.g.s in left-hand image) should be included.	71
Figure 7-4. Putamen protocol example: ensure that edges are smooth and biologically plausible in both views.	71
Figure 7-5. A) Raw putamen volume estimates in each group separated by scanner types; B) Raw putamen volume estimates in each group separated by scanner types and subtypes.	74
Figure 7-6. Bland Altman (left) and scatter plot (right) comparisons of automated longitudinal putamen volume change estimates (ml) against the manual gold-standard. Scatters of control (●) and HD (x) data-points between the manual volume change estimate (ml; y-axis) and the automated estimates of change (x-axis) also show the line-of-equality.	87
Figure 7-7. Bland Altman and scatter plot comparisons of automated longitudinal putamen volume change estimates (ml) against the CBSI gold-standard. Scatters of control (●) and HD (x) data-points between the CBSI volume change estimate (ml; y-axis) and the automated estimates of change (x-axis) also show the line-of-equality.	92
Figure 8-1. Cerebellum delineation protocol development examples:	98
Figure 8-2. a) Unadjusted cerebellar volumes in control and HD groups scanned on Philips and Siemens scanners with lower thresholds of 65% (blue) and 70% (red) respectively; b) The differences in cerebellar volume outputted from 65% and 70% lower thresholds across groups and scanner types; c) Volumes outputted from both scanner types utilising the scanner-specific lower thresholds (65% for Philips and 70% for Siemens).....	101
Figure 8-3. A) An example of a participant whose serial scans (left = baseline, right = follow-up) show severe geometric distortion most obviously affecting the chin, neck and skull (highlighted by red arrows). B) A fluid VCM between two scans showing extreme cerebellar enlargement due to distortions of the cerebellar and brainstem region.	104
Figure 8-4. Examples of cerebellum segmentations from the two automated software methods in this method comparison:	107
Figure 9-1. An illustration of two modelled datasets: in ‘A’ age is related to cortical thickness in a consistent way between the four study groups; in ‘B’ age effects the groups in different ways. If the data looks like ‘A’ then the DOSS model is valid. If the data looks like ‘B’ the DODS model must be used (difference slopes between groups and genders as age increases).....	113
Figure 9-2. Scatter and Bland Altman plots comparing PCB and SPC estimates in whole-cortex thickness change.	117
Figure 9-3. Significance maps of cross-sectional between-group differences in cortical thickness analysed with both a 10mm (left) and 20mm (right) smoothing kernel.	120
Figure 9-4. Magnitude maps of between-group differences in rate of change over 6-, 9- and 15-month intervals analysed with both a 10mm (left) and 20mm (right) smoothing kernel.	121
Figure 9-5. Magnitude (left) and significance maps, with (middle) and without (right) FDR adjustment (adj.) for multiple comparisons, of atrophy over six, nine and 15 months in HD participants - compared to zero.	123
Figure 9-6. Cortical thickness averages (mm) within each lobe at baseline as outputted by both FreeSurfer versions 5.1.0 and 5.3.0.....	125
Figure 9-7. Rate of thickness change (mm) over 15 months within each lobe as outputted by both FreeSurfer versions 5.1.0 and 5.3.0.....	125
Figure 9-8. (Left) Magnitude, (middle) variance and (right) significance maps of between-group differences within the left and right hemispheres in: A) cross-sectional thickness (mm); B) rate of change over six	

months C) nine months and; D) 15 months. All analyses are adjusted for age, gender, study site and scan interval. Variance is calculated as the square of the standard error. The significance map colour bars represent the T values between $p < 0.001$ and $p < 0.05$ adjusted for multiple comparisons using FDR correction ($p < 0.05$).	132
Figure 10-1. Examples of static photo cues expressing happiness, anger, disgust and fear (top-bottom, left-right).	140
Figure 10-2. Estimated between-group differences (HD vs control) in number of errors out of five for each emotion modality combination, with 95% bootstrapped bias-corrected and accelerated confidence intervals. Positive values indicate that HD participants made more errors than controls. Statistical significance is highlighted with: * $p < 0.05$; ** $p < 0.01$.	144
Figure 13-1. Examples of: A) a volumetric cerebellum delineation in MIDAS software; B) GM (yellow) and WM (white) masks for the diffusion analysis overlaid on the corresponding average B0 image.	156
Figure 13-2. A statistical parametric map showing t scores of significant ($p < 0.05$) differences between controls and early HD patients in cerebellar WM volume, adjusted for age, gender, study site, TIV and alcohol intake history. Results are adjusted for multiple comparisons using FDR correction ($p < 0.05$).	159
Figure 13-3. Plots of raw HD group data for the significant and borderline associations between clinical scores and imaging metrics.	161
Figure 13-4. Associations between cerebellar volume and HADS-SIS plotted for both the HD (red (x)) and control (blue (•)) groups with the unadjusted regression lines of best fit.	162
Figure 14-1. WM tracts in which FA significantly associated with recognition of anger, disgust, fear, happiness, sadness and surprise (top to bottom) in the HD group. Colours are random for each region and do not represent any strength to the observed associations.	175
Figure 15-1. Occipital cortex thickness (mm) plotted by subregion and subgroup. The box plot whiskers range from the minimum to the maximum values (excluding outliers), the box spans from the 25 th to the 75 th percentile and the central line represents the median value.	184
Figure 15-2. A. The LOC, cuneus, lingual and pericalcarine atlas regions as defined by the Desikan-Killiany atlas overlaid on the study-specific average cortical template and on the inflated study-specific average cortical template. B. Significance maps of the associations ($0.0001 < p < 0.05$) between occipital cortex thickness and cognitive task performance displayed on inflated brain templates. Associations were adjusted for age, gender, study site, education, CAG, disease burden score and frontal lobe thickness, and corrected for multiple comparisons using Monte Carlo cluster-wise correction ($p < 0.05$) across the four occipital regions.	187
Figure 18-1. Cross-sectional effect size estimates (mean and 95% CIs) for each of the imaging outcomes. Data are grouped by imaging metric and modality. Effect sizes were calculated as the absolute difference in the mean of the metric between groups, adjusted for age, gender and study site, divided by the estimated residual SD of the HD group.	202
Figure 18-2. Regions of significant 6- (left) and 9-month (right) WM atrophy in the HD group compared with controls, adjusted for multiple comparisons using FWE at $p < 0.05$. The colour bars represent the T-scores.	208
Figure 18-3. Guidelines for biomarker selection over short and varying time intervals in future clinical trials of potentially disease-modifying therapies for early stage HD. Sample size requirements are per treatment arm; calculated using the standard formula (Julious 2009), with 90% power and two-tailed $p < 0.05$, for therapies with 20% and 50% estimated treatment efficacy in early-stage HD.	210
Figure 23-1. a) Several seeds may be needed to roughly fill the structure on each slice; b) Manually edit the lateral border, disconnecting the putamen from the WM and claustrum.	226

Figure 23-2. Two examples of the most inferior slice in which the putamen should be segmented (left) – here the putamen is still clearly separated by WM from the caudate (red arrows). Once this separation is no longer clear (e.g. right images) stop the segmentation and do not segment this slice.	226
Figure 23-3. If the tail and head of the putamen are connected on that slice segment the whole structure (left). When this becomes split by WM deseed the tail (right).	227
Figure 23-4. Ensure that edges are smooth and biologically plausible in both views.	227
Figure 23-5. A) An example of a starting slice. Seed the superior anterior regions of cerebellum. B) Cut around the top edge of the cerebellum removing the cerebrum and middle to remove all brainstem.	229
Figure 23-6. Highlights nerves that may be confused for cerebellar GM but should not be included in the segmentations.	229
Figure 23-7. An example of WM removal before the hemispheres join. Note the seeded vermis between the hemispheres. The same segmentation is shown in the coronal and sagittal views.	230
Figure 23-8. When the GM of the hemispheres become connected by the vermis, include this and the underlying WM arch. Continue the curve of the connective WM down to the inferior folia (yellow line is correct, the red segmentation line is wrong).	230
Figure 23-9. Include central GM (red arrows), only removing the WM of the brainstem.	231
Figure 23-10. Remove the connective tissue between the posterior lobes.	231
Figure 23-11. Example of a cerebellum segmentation in MIDAS software.	232

Table of Tables

Table 2-1. Summary table of longitudinal observational studies in HD reporting caudate and putamen volume change over time.	26
Table 2-2. Summary table of longitudinal observational MRI studies in HD reporting whole-brain volume change over time.	28
Table 2-3. Summary table of longitudinal observational MRI studies in HD reporting GM and WM volume change over time.	31
Table 2-4. A summary of quantitative measures of diffusion MR images.	32
Table 2-5. A summary table of the longitudinal DTI studies in HD.	33
Table 2-6. Details of clinical trials in HD to-date utilising neuroimaging biomarkers.	36
Table 3-1. PADDINGTON study participant demographics – adapted from Hobbs et al. (Hobbs et al. 2013).	39
Table 7-1. Between-scanner test of the putamen segmentation protocol - sample demographics.	73
Table 7-2. Demographics of putamen and caudate method comparison samples.	76
Table 7-3. A systematic visual assessment and summary of putamen segmentation errors outputted from different automated methods.	80
Table 7-4. Summary statistics of raw putamen 6- and 15-month volume change data.	84
Table 7-5. 6- and 15-month percentage change in putamen volumes, with adjusted between-group differences, p-values and effect sizes.	88
Table 7-6. Summary statistics of raw caudate 6- and 15-month volume change data.	90
Table 7-7. 6- and 15-month percentage change in caudate volumes, with adjusted between-group differences, p-values and effect sizes.	93
Table 8-1. Cerebellum lower threshold comparison participant demographics.	100
Table 8-2. Cerebellum method comparison participant demographics.	106
Table 8-3. Summary table of cross-sectional cerebellum volume measurements (ml), with adjusted between-group differences and p-values.	108
Table 8-4. Summary table of longitudinal cerebellum volume change estimates over 24 months (ml; adjusted for interval), with adjusted between-group differences & p-values.	108
Table 8-5. Spearman’s rank correlation coefficients for the associations between 24-month cerebellar change estimates (ml). Correlations with the 24-month BBSI (ml) are also included.	109
Table 9-1. Demographics of the PADDINGTON cohort with FreeSurfer analysed scans at both baseline and 15-month visits.	114
Table 9-2. Rate and percentage estimates of cortical thickness change over 15 months (adjusted for interval) in averaged lobular regions, with adjusted between-group differences and effect sizes.	116
Table 9-3. Rate of lobular cortical thickness change over 15 months outputted by FreeSurfer versions 5.1.0 and 5.3.0, with adjusted between-group differences and effect sizes.	127
Table 9-4. Rate of cortical thickness change (mm) over 15 months (adjusted for interval) extracted from all atlas regions, with adjusted between-group differences and effect sizes.	134
Table 10-1. Emotion recognition task participant demographics.	142
Table 10-2. Mean (SD) number of errors in both the control and HD groups for: each emotion within each stimulus modality (/5; e.g. photos of fear); each emotion across all stimulus modalities (/15; e.g. fear recognition from photos, vocal and film cues); total within each stimulus modality across all emotions (/30; e.g. total number of errors from all photo stimuli); and the total overall (/90).	143

Table 10-3. A matrix of the percentage of answers given by the HD group for each emotion displayed, separated by stimulus modality.	146
Table 13-1. Cerebellum imaging investigation participant demographics and characteristics.	158
Table 13-2. Summary statistics and adjusted between-group differences in cerebellar imaging metrics... ..	158
Table 13-3. Imaging associations with clinical scores in the HD group (n=22) – association coefficient (95% CI) p-value.....	160
Table 14-1. Macro-structural imaging associations with emotion recognition errors within stimulus modalities.	167
Table 14-2. Macro-structural imaging associations with emotion-specific recognition errors.	168
Table 14-3. Micro-structural FA associations, across all JHU-atlas regions tested, with emotion recognition errors within stimulus modalities. Estimates (95% CI) and p-values are adjusted for age, gender, BFRT score and motor response time.	170
Table 14-4. Micro-structural FA associations, across all JHU-atlas regions tested, with emotion-specific recognition errors. Estimates (95% CI) and p-values are adjusted for age, gender, BFRT score and motor response time.	172
Table 15-1. Demographic data. Data are presented as mean (SD) range.	183
Table 15-2. Adjusted between-group differences in occipital cortical thickness measures; reported with p-values and effect sizes.	185
Table 15-3. Associations, in the HD group, between cortical thickness within occipital regions and cognitive and motor task performance. Light grey highlights statistically significant ($p < 0.05$) associations.....	186
Table 17-1. 24-month effect sizes (95% CI) from the TRACK-HD study (Tabrizi et al. 2012). All estimates are adjusted for age, sex, education level and study site with the exception of the neuroimaging measures, which were adjusted for age, sex and study site only. CV = coefficient of variation. PBA = problem behaviours assessment. * As controls and premanifest HD groups are not expected to show any change in UHDRS functional capacity, this outcome was modelled and effect sizes are presented for symptomatic HD groups only. When calculating effect sizes, change over 24 months in the HD groups was compared with zero expected change rather than estimated change in controls.....	192
Table 17-2. PREDICT-HD effect sizes for two-year volume change (Aylward et al. 2011). * N per treatment arm for a 2-year clinical trial, assuming two-sided $p = 0.05$; 90% power. † % reduction in rate of case-control atrophy difference.....	193
Table 17-3. PREDICT-HD 10-year data (Paulsen et al. 2014): Estimated sample size requirements (right side) for a two-arm phase II randomised clinical trial. Effect size is the percentage difference in rate of change (slope) of the treated and untreated groups. SP-Tapping, speeded tapping; Brady, bradykinesia.....	194
Table 18-1. Biomarker group averages at baseline, 6- and 15-month visits, with adjusted between-group differences in change estimates.....	204
Table 18-2. 6-, 9- and 15-month regional BSI and fluid-based estimates of change with adjusted between-group differences and p-values.	206
Table 18-3. 6-, 9- and 15-month effect size estimates from the PADDINGTON study.	207

List of Abbreviations

3T	3 Tesla
AD	Axial diffusivity
AIR	Automated image registration
BBSI	Brain boundary shift integral
BFRT	Benton Facial Recognition Test
BOLD	Blood oxygenation level dependent
BRAINS	Brain Research: Analysis of Images, Networks and Systems
BSI	Boundary shift integral
CAG	Cytosine adenine thymine, the codon that encodes the amino acid glutamine
CAP	CAG-Age Product
CBSI	Caudate boundary shift integral
CC	Corpus callosum
CI	Confidence interval
CNS	Central nervous system
CSF	Cerebrospinal fluid
Ctls	Controls
DARTEL	Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra
DoF	Degrees of freedom
DTI	Diffusion tensor imaging
ES	Effect size
FA	Fractional Anisotropy
FDR	False discovery rate
FIRST	FMRIB's integrated registration and segmentation tool
fMRI	Functional magnetic resonance imaging
FOV	Field of view
FSL	FMRIB software library
FWE	Family-wise error
FWHM	Full width at half maximum
GM	Grey matter
HADS-SIS	A combined psychiatric assessment comprised of components from the Hospital Anxiety and Depression Scale and the Snaith Irritability Self-assessment scale
HD	Huntington's Disease
Htt	Huntingtin protein
HVLT	Hopkins verbal learning test
LOC	Lateral occipital cortex
MD	Mean diffusivity
MIDAS	Medical information display and analysis system
MNI	Montreal Neurological Institute
MRI	Magnetic resonance imaging
MRS	Magnetic Resonance Spectroscopy
N	Number
Ns	Not significant
PADDINGTON	Pharmacodynamic approaches to demonstration of disease-modification in Huntington's Disease by SEN0014196
PCB	Percentage change from baseline

PDD	Primary diffusion direction
PET	Positron emission tomography
PreHD	Premanifest Huntington's Disease
PVE	Partial volume effect
QC	Quality control
QNE	Quantified Neurological Exam
RD	Radial diffusivity
ROI	Region of interest
S	Subject/Participant
SD	Standard deviation
SDMT	Symbol Digit Modalities Test
SFOF	Superior fronto-occipital fasciculus
Sig.	Significance
SOP	Standard operational procedure
SPC	Symmetrized percent change
SPM	Statistical parametric map
STEPS	Similarity and Truth Estimation for Propagated Segmentations
STG	Superior temporal gyrus
TBSS	Tract-based spatial statistics
TFC	Total Functional Capacity
TIV	Total intracranial volume
TMS	Total Motor Score
TPM	Tissue probability map
UHDRS	Unified Huntington's Disease Rating Scale
VBM	Voxel-based morphometry
VCM	Voxel compression map
WM	White matter
WP2	Work package 2
Yr	Year
YTO	Years to onset

Aims of this Thesis

At this stage of HD research the optimisation and automation of biomarkers to assess therapeutic efficacy in HD is a major priority. Consequently, the overall aims of this thesis (addressed in Parts 1-3) are to:

1. Develop and evaluate tools sensitive to neurodegenerative change.
2. Perform exploratory investigations of neuroimaging associations with clinical and cognitive symptoms to further enhance our knowledge of HD and strengthen the argument for the clinical relevance of neuroimaging as a surrogate marker of disease manifestation and progression.
3. Conduct a direct statistical comparison of a comprehensive battery of biomarkers in HD and subsequently produce guidelines for future application in clinical trials.

1 Background: Huntington's Disease (HD)

This introductory chapter will describe the underlying genetic cause of HD and the resultant neuropathology and clinical phenotype, as well as the scope of present-day treatment options. This will be followed by an introduction to current research methods in HD; clinical rating scales, biomarkers, participant cohorts and finally neuroimaging study set-up.

1.1 Genetics

HD is a genetic neurodegenerative disease that affects approximately 6-7 per 100,000 people in the UK, although more recent reports suggest the incidence could be more than double this estimate (Wise 2010). HD is caused by an autosomal dominant CAG trinucleotide repeat expansion in exon 1 of the IT15 (interesting transcript 15) gene. Each C-A-G sequence codes for the amino acid glutamine. The IT15 gene codes for the protein huntingtin (Htt) which was identified in 1993 as a result of an international collaborative effort (Huntington's Disease Collaborative Research Group 1993).

The Htt gene varies in length depending on the number of CAG repeats it contains: CAG repeats of 6-35 are within the normal, non-pathological range; CAG repeats of 40+ show full penetrance (i.e. 100% of individuals are destined to develop HD); whereas incomplete penetrance is observed in the intermediate range of 35 to 39 repeats. Intermediate repeat number allele carriers, even if symptom free over their life-time, can have children with CAG repeats of over 40 as a result of instability during spermatogenesis (Ranen et al. 1995).

Genetic testing is available for at-risk individuals (e.g. those with a parent with HD) over the age of 18. In Europe (Morrison 2010;Morrison et al. 2011) and Australia (Tassicker et al. 2009) approximately 12-15% of people at-risk of carrying the HD gene choose to find out their gene-status. This is reduced to just ~10% in the US and Canada (Lancet Editorial 2010), most likely affected by employment and insurance worries. The genetic test confers on HD research the advantage of foresight and the ability to identify, with 100% confidence, individuals who will develop the disease in later life but who may not have any manifest symptoms or signs of disease. These individuals are described as pre-symptomatic and referred to in this thesis as preHD cohorts.

As well as determining disease status, the CAG repeat length is also thought to influence 50-70% of the age of disease onset (Andrew et al. 1993;Brinkman et al. 1997) and potentially also the rate of disease progression (Rosenblatt et al. 2006;Rosenblatt et al. 2011;Tabrizi et al. 2013), although findings differ on this point (Kiebert et al. 1994).

1.2 Neuropathology

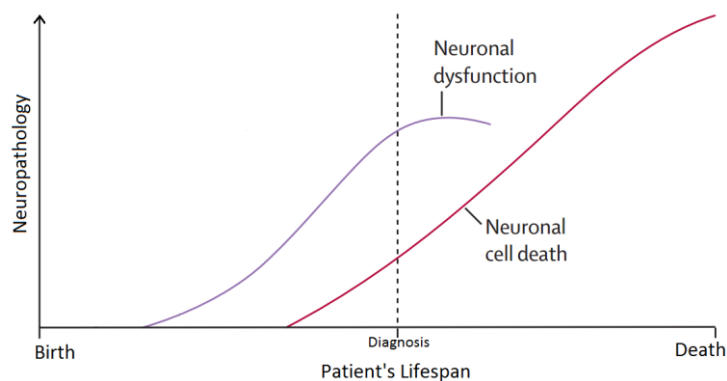


Figure 1-1. A model of the progression of HD neuropathology over a patient's lifespan from preHD through diagnosis (represented as a dotted line) to manifest disease. Adapted from Ross & Tabrizi (Ross and Tabrizi 2011).

Neuropathology has been detected in individuals estimated to be over a decade from disease onset (Paulsen et al. 2008). Neuronal dysfunction is thought to precede cell death. Cell death and consequently brain atrophy becomes more severe and widespread with disease progression (Aylward et al. 2000). This process is modelled in a simplified format in Figure 1-1.

Post-mortem studies have localised the most severe HD neuropathology to the GABAergic medium spiny neurons within the caudate and putamen of the striatum, but this spreads throughout the brain with disease progression (Halliday et al. 1998; Vonsattel et al. 1985; Vonsattel and DiFiglia 1998). By end-stage disease almost the entire cortical mantle, substantial subcortical grey matter (GM) and widespread white matter (WM) tracts are affected, with whole-brain volume reduced by up to 30% (de la Monte et al. 1988). With the emergence of neuroimaging technology it has become possible to track disease-related brain changes *in vivo*. Chapter 2 reviews the neuroimaging literature in HD.

1.3 Clinical Presentation

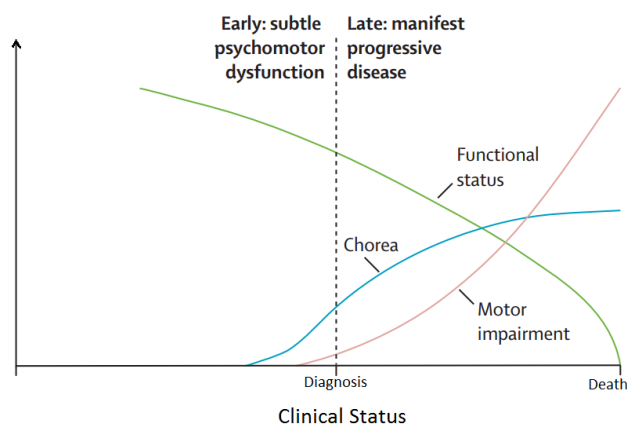


Figure 1-2. A model of the progression of HD symptoms over a patient's lifespan. Adapted from Ross & Tabrizi (Ross & Tabrizi 2011).

Subtle psychomotor dysfunction is reported in preHD cohorts many years before formal diagnosis (Hart et al. 2011), depicted in Figure 1-2 as a dotted line. Diagnosis is based on the emergence of motor signs, assessed and rated by a clinician, and judged to be unequivocal signs of HD. This typically occurs between 35 and 50 years of age. Onset is a gradual process and therefore formal diagnosis is subject to inter-rater variability (de Boo et al. 1998). Death typically occurs between 15-20 years post-onset.

HD manifests with a triad of symptoms: motor, cognitive and psychiatric (Craufurd and Snowden 2002; Novak and Tabrizi 2011). Symptoms and signs in HD are heterogeneous and vary as the disease progresses. In early-stage disease minor motor abnormalities may be present: restlessness, abnormal eye movements, hyper-reflexia, impaired finger tapping, fidgety movements of fingers, hands and toes. These choreic symptoms then develop into a more rigid, functionally-disabling motor impairment: dystonia (abnormal muscle tone resulting in muscular spasm and abnormal posture), bradykinesia (slowness of movement) and incoordination. Neuropsychiatric symptoms are common, including depression, personality changes and psychosis. This can involve explosive temper, aggression, compulsions, hyper-sexuality and violence. Cognitive impairment is a near-universal feature of early HD but varies in severity. In HD this is not a global dementia but rather a psychomotor slowing: executive dysfunction, distractibility, apathy and memory deficits.

1.4 Treatment Scope

HD is currently incurable although symptomatic treatments are available, most commonly alleviating the symptoms of chorea (e.g. Tetrabenazine) and psychiatric symptoms (selective serotonin reuptake inhibitors and neuroleptics) (Ross & Tabrizi 2011). Several potentially curative therapies are however in varying stages of development (Handley et al. 2006; Maucksch et al. 2013; Ross & Tabrizi 2011; Zhang and Friedlander 2011). To-date only a few have made it into clinical trials in HD, none of which demonstrated a significant therapeutic benefit.

Of the potentially symptomatic treatments of HD there have been trials of Dimebon (HORIZON Investigators of the Huntington Study Group and European Huntington's Disease Network. 2013; Kiebert et al. 2010), Memantine (Hjermind et al. 2011) (Mejia et al. 2005), Tetrabenazine (Jankovic et al. 2004), Pridopidine (Lundin et al. 2010) and SD-809 (NCT01897896 & NCT01795859). Of the potentially disease-modifying treatments trials of the following have been, or are being, run: Ethyl-EPA (Huntington Study Group TREND-HD Investigators 2008), Minocycline (Huntington Study Group DOMINO Investigators 2010), Co-enzyme Q10 (Hyson et al. 2010), Creatine (Rosas et al. 2014), Selisistat (the PADDINGTON study; NCT01521585) and Lamotrigine (Kremer et al. 1999). However, recent data suggest that these trials were not sufficiently powered (Tabrizi et al. 2012). Larger trials are forecast in the near future.

1.5 Research Methods

1.5.1 Clinical Rating Scales

Unified HD Rating Scale (UHDRS)

The Unified HD Rating Scale (UHDRS (Anon 1996)) is a clinical rating tool which includes: motor, functional, cognitive, behavioural and emotional components. This thesis reports Total Motor Score (TMS) and Total Functional Capacity (TFC) – scale items are detailed in Appendix Sections 23.1 and 23.2.

The TMS is scored within the range of 0 (no motor signs) to 124 (severe motor impairment) and includes ratings of ocular pursuit, saccade initiation and velocity, dysarthria (difficult or unclear articulation of speech), tongue protrusion, finger tapping, luria (fist-hand-palm test), rigidity, bradykinesia, dystonia, chorea, gait, tandem walking and the retropulsion-pull and pronate/supinate-hand tests. It also includes a diagnostic confidence score: a grading from 0-4 from normal to ≥99% confidence in the presence of motor abnormalities that are unequivocal signs of HD. Official diagnosis requires a confidence score of four.

TFC is scored on a scale of 0 (complete dependence) to 13 (full capacity) and covers the patient's care level requirement and ability to maintain: an occupation, finances, domestic chores and activities of daily living (eating, dressing, bathing). TFC is used to define the stages of disease – details in Section 1.5.3.

Years to Onset (YTO)

PreHD is a heterogeneous group label and it is therefore useful to be able to classify how far each participant is from estimated disease onset. Since there is such a direct relationship between CAG repeat length and age at onset, this gives us a unique opportunity to try to predict the number of years to onset (YTO) and therefore more accurately map the premanifest disease course. For a review of the available approaches see Langbehn et al. (Langbehn et al. 2010). The Aylward (Aylward et al. 1996) and Langbehn (Langbehn et al. 2004) algorithms are arguably the most widely used methods for estimating YTO in HD.

The algorithm published by Langbehn et al. is based on data from 2913 pre- and manifest patients from 40 centres worldwide. They used probabilistic modelling (a non-linear parametric survival analysis) to describe the relationship between CAG and current age and consequently predict the probability of disease-onset at different ages for different patients. For example, the model predicts a 91% chance that a 40-year-old individual with 42 repeats will have onset by the age of 65, with a 95% confidence interval (CI) from 90 to 93%. Estimates are typically reported in the literature based on a 60% likelihood of onset.

Aylward et al. derived a prediction equation from stepwise multiple regression analysis of data from a sample of 50 symptomatic parent-child pairs:

$$\text{Age at onset} = (-0.81 \times \text{CAG repeat length}) + (0.51 \times \text{parental onset})$$

These two methods can produce markedly different estimates for the same preclinical cohort, e.g. Majid et al. (Majid et al. 2011a). Inconsistency and inaccuracies are most likely due to lack of sensitivity to additional influential factors such as: environment (Wexler et al. 2004), genetic interactions (Li et al. 2003), paternal versus maternal transmission (Ranen et al. 1995) and interactions between the expanded and normal allele (Djousse et al. 2003).

Disease Burden

Disease burden score can be interpreted as an index of an individual's lifetime exposure to mutant Htt toxicity, and consequently disease severity. This rating scale was developed by Penney et al. (Penney et al. 1997). Autopsy analysis of 89 HD brains found a linear correlation between the CAG repeat number and the quotient of the degree of atrophy in the striatum divided by age at death, with an intercept at 35.5 repeats:

$$\text{Striatal atrophy} = (\text{CAG} - 35.5) \times \text{age at death}$$

The largest CAG repeat length, therefore, at which no pathology is expected to develop, is 35.5. These results imply that striatal damage in HD is almost entirely a linear function of the length of the CAG repeat length beyond 35.5 repeats multiplied by the age of the patient. Thus, it is predicted that the pathological process develops linearly from birth. This formula has been adapted to provide estimations of disease burden during the disease process:

$$\text{Disease burden} = (\text{CAG} - 35.5) \times \text{current age}$$

Disease burden scores span the whole spectrum of pre- to manifest-HD, providing a continuous scale. This measure however should be used with caution for the following reasons: the formula was developed at post-mortem and extrapolation to earlier stages of the disease and across time-points based on cross-sectional data may therefore introduce errors; data is based on striatal atrophy whereas it is known that widespread brain regions and other factors such as environment (Wexler et al. 2004) and genetics (Li et al. 2003) also affect the severity of disease expression; and the pathological process may not develop linearly (an assumption of this model).

Many studies are now moving towards the use of a score called the CAG-Age Product (CAP (Zhang et al. 2011)). This score is derived from a standard parametric survival model based on data from 730 preHD individuals and defined as follows:

$$\text{CAP} = 100 \times \text{current age} \times [(\text{CAG} - L) / S]$$

S is a normalizing constant chosen so that the CAP score is approximately 100 at the patient's expected age of onset as estimated by Langbehn et al. (Langbehn et al. 2004). L is a scaling constant that anchors CAG length approximately at the lower end of the distribution relevant to HD pathology; calculated as 35.5 by Penney et al. (Penney et al. 1997) but estimates vary (Warner and Hayden 2012; Zhang et al. 2011).

1.5.2 Biomarkers

With the commencement of disease-modifying clinical trials in humans comes the need for reliable and sensitive biomarkers of disease progression which may prove useful as outcome measures. The Biomarkers Definitions Working Group (Atkinson et al. 2001) defines a biomarker as,

‘A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention’.

Characteristics of an ideal biomarker include quantification which is reliable, reproducible across sites, minimally invasive and widely available. The biomarker should show low variability in the normal population and change linearly with disease progression, ideally over short time intervals. Finally, the biomarker should respond predictably to an intervention which modifies the disease.

Biomarkers fall into two categories: those that provide clinical end-points and those that provide surrogate end-points. Clinical end-points are those which directly measure a patient’s experience, including level of functioning, quality of life and survival. Surrogate end-points are expected to be related to clinical benefit, for example metrics from neuroimaging and biochemical biomarkers.

The clinical, neuroimaging and biochemical biomarkers developed for HD have been reviewed by Weir et al. (Weir et al. 2011) and Andre et al. (Andre et al. 2014). Clinical biomarkers, involving scales of symptom severity, have been most widely used as the primary outcome measures in the first clinical trials in HD. These include assessments from the UHDRS (1996) and Quantified Neurological Exam (QNE; (Folstein et al. 1983)). Unfortunately, inter-rater variability (de Boo et al. 1998), floor and ceiling effects (Mickes et al. 2010), low sensitivity to longitudinal change (Tabrizi et al. 2011) and inability to easily discriminate between disease modification and symptomatic benefit are all potential limitations of these measures. Such limitations mean that well-powered human clinical trials based on these measures would be unfeasibly large and prohibitively expensive. In fact, as previously mentioned, evidence suggests that all trials in HD to-date have been under-powered (Tabrizi et al. 2012). Neuroimaging biomarkers are obvious candidates as additional outcome measures in future large-scale clinical trials because of their clear relevance to the neuropathology of disease and their increased precision and sensitivity compared with some standard functional measures (Tabrizi et al. 2011).

1.5.3 Research Cohorts

Recruitment of patients to HD studies is hindered by its rarity and the requirement, in some trials, for participants to have had a positive genetic test. Sample size limitations can be overcome with large multi-site studies, for example PREDICT-HD (n=1314 (Aylward et al. 2011;Paulsen et al. 2006;Paulsen et al. 2014)) and TRACK-HD (n=366 (Tabrizi et al. 2009)), but with this come additional logistical issues - discussed in Section 1.5.4. In addition to sample size, it is important that the cohorts are well-characterised in terms of

age, gender, IQ, socioeconomic status, education level and CAG repeat lengths since these factors are known to influence disease onset and progression in HD (Ross & Tabrizi 2011).

Control groups enable us to establish baselines and rates of change in normal ageing (Sullivan and Pfefferbaum 2007). In order to match patients to controls as closely as possible in terms of age, educational level, background, home life etc. many HD studies recruit partners, spouses and gene-negative siblings as controls.

HD cohorts can be categorised as preHD, early-stage HD (Stages I (TFC 13-11) and II (TFC 10-7)) or late-stage HD (Stages III (TFC 6-3) and IV (TFC 2-1)) (Shoulson and Fahn 1979). The staging after diagnosis is based on TFC but some studies prefer to use duration of illness, disease burden score or CAP score. These clinical stages can be argued to be somewhat arbitrary cut-offs when applied to imaging studies in which atrophy can be measured on a continuous scale. Nevertheless, there is evidence that atrophy measures have strong predictive power for clinical conversion to manifest disease (Tabrizi et al. 2013).

1.5.4 Neuroimaging Study Set-Up

The previous sections have discussed research methods generalisable to all studies in HD. There are however important additional considerations during the set-up phase for neuroimaging studies and trials. TRACK-HD and PREDICT-HD, mentioned previously, are large multi-site observational studies utilising longitudinal MR imaging which were designed to imitate clinical trial set-ups. The challenges associated with large-scale organisation, quality control (QC) and assurance, data management, anonymization, storage, analysis and data dissemination have been addressed by these studies. Inter-scanner differences and consistency of acquisition protocol have been highlighted as important areas to consider. Motion artefacts, a particular problem in manifest (choreic) HD, should be minimised by consideration of disease progression during the course of the study, as well as employing a quality assurance procedure during scan acquisition and thorough QC of all acquired scans, with rescans obtained where necessary. There is some evidence that differing field strengths create volume difference bias (Jovicich et al. 2009) and therefore this should also be a consideration.

2 Neuroimaging in HD: A Literature Review

Many observational studies and reviews have examined neuroimaging measures cross-sectionally in HD, comparing different stages of illness to make assumptions about disease progression. Fewer however have acquired serial scans and assessed longitudinal change directly; for a review see Rees et al. (Rees et al. 2013). It is these findings that are most relevant when considering imaging biomarkers as outcome measures for clinical trials of disease-slowing compounds and therefore constitute the focus of this literature review.

2.1 Structural MRI

Structural magnetic resonance imaging (MRI) allows assessment of the macro-structural effects of the underlying neuropathology of HD, namely brain atrophy. MR volumetry is well developed and widely published in HD (Georgiou-Karistianis et al. 2013b). The findings from the main regions of interest (ROIs) in HD are outlined below – MR image analysis methods used in these studies are described in Section 4.2.

2.1.1 Basal Ganglia

Since the most striking pathological changes in HD are found in the basal ganglia (Halliday et al. 1998) many MRI studies have focused on this region. Applying manual delineation, Aylward et al. were the first to demonstrate longitudinal atrophy of this structure in manifest (Aylward et al. 1997) and preHD (Aylward et al. 2000). These findings have been replicated in multiple studies (Table 2-1). Over a 24-month period TRACK-HD found there to be significantly greater atrophy within the caudate of progressors (preHD participants with an increase in TMS of five points or more, any TFC decline, or a new diagnostic confidence score of four) than non-progressors (Tabrizi et al. 2012). There was no significant difference between progressors and non-progressors in putamen atrophy rate. This suggests that caudate atrophy is associated, directly or indirectly, with one or all of these clinical measures. Over 36 months the TRACK-HD study reported the largest effect sizes, compared with controls, in the caudate over all other neuroimaging measures (Tabrizi et al. 2013).

Study Reference	Cohort	Interval (Months)	YTO/Years Duration (D)	Method	Caudate Volume Loss	Statistical Sig.	Putamen Volume Loss	Statistical Sig.
	n	Mean (SD)	Mean (SD)		Mean (SD)		Mean (SD)	
(Aylward et al. 1997)	23 HD	20.8 (7.33)	7.4 (4.01) D	Manual	9.5 %	sig.>0	6%	sig.>0
(Aylward et al. 2000)	10 preHD	35.8 (12.9)	2.04 (3.77) YTO	Manual	5.48 (5.23) %	sig.>0		
	10 mild HD		8.1 (5.4) D		17.16 (14.45) %	sig.>0		
	10 mod HD		9.8 (3.3) D		22.13 (13.74) %	sig.>0		
(Aylward et al. 2003)	19 HD	29.5 (3.6)	Stage I or II * Clinical trial	Manual	0.64 (0.37) ml	sig.>0		
(Aylward et al. 2004)	17 preHD	50.4 (21.6) *2-6 scans	5.4 (5.5) YTO	Manual	4.3 %/yr	sig.>0	3.1 %/yr	sig.>0
(Hobbs et al. 2009)	17 preHD	27.6 (1.2)	17 (6.2) YTO	Manual or CBSI (Hobbs et al. 2009)	Manual: 1.7 (1.2) %/yr CBSI: 1.2 (0.9) %/yr	sig.>ctls sig.>ctls		
	26 HD	27.6 (1.2)	4.9 (2.8) D		Manual: 3.4 (1.7) %/yr CBSI: 2.9 (1.6) %/yr	sig.>ctls sig.>ctls		
	13 ctls	26.4 (1.2) *all with 3 annual scans			Manual: 0.2 (1.0) %/yr CBSI: 0.1 (1.0) %/yr			
(Vandenberghe et al. 2009)	8 HD	25 (3.1)	1.5 (1.2) D	Manual	5.70 (3.56) %/yr	sig.>0	3.51 (2.26) %/yr	sig.>0
(Hobbs et al. 2010a)	40 HD	12 & 27	Stages I & II	Manual	0.158 ml/yr	sig.>0 & ctls		
	19 ctls	*2 or 3 scans			0.013 ml/yr	ns>0		
PREDICT-HD (Aylward et al. 2011)	82 preHD (far)	24.4 (1.6)	>15 YTO	BRAINS2 (Magnotta et al. 2002)	0.32 (0.41) ml	sig.>ctls	0.25 (0.43) ml	sig.>ctls
	73 preHD (mid)	24.4 (1.4)	9-15 YTO		0.48 (0.49) ml	sig.>ctls	0.42 (0.43) ml	sig.>ctls
	56 preHD (near)	24.6 (1.9)	<9 YTO		0.36 (0.41) ml	sig.>ctls	0.4 (0.45) ml	sig.>ctls
	60 ctls	24.1 (1.1)			0.12 (0.4) ml		0.13 (0.56) ml	
(Majid et al. 2011a)	36 preHD	12 (1.2)	6.3 (7.3) YTO (Aylward), 14.4 (7.2) YTO (Langbehn)	Quarc (Holland et al. 2009;Holland and Dale 2011)	1.58 (1.42) %/yr	sig.>ctls	0.95 (0.68) %/yr	sig.>ctls
	22 ctls				0.37 (1.31) %/yr		0.19 (0.69) %/yr	
TRACK-HD (Tabrizi et al. 2011)	116 preHD	11.5 (0.8)	10.8 YTO	CBSI & BRAINS2	1.37 (0.99 to 1.75) %/yr	sig.>ctls	2.25 (1.29 to 3.21) %/yr	sig.>ctls
	114 HD	11.6 (0.8)	Stages I & II		2.86 (2.34 to 3.39) %/yr	sig.>ctls	4.45 (2.94 to 5.96) %/yr	sig.>ctls
	115 ctls	11.6 (0.8)			0.62 %/yr		1.17 %/yr	
TRACK-HD (Tabrizi et al. 2012)	33 preHD prog.	24	10.8 YTO	CBSI & BRAINS2	5.589 (2.231) %	sig. diff. between groups	7.174 (5.148) %	ns diff. between groups
	78 preHD non-prog.				3.56 (2.407) %		5.339 (3.620) %	
IMAGE-HD (Dominguez et al. 2013)	31 preHD	18	14.7(8) YTO	Semi-customised SPM8	1.92 (SE 0.58) %	sig.>ctls	0.46 (SE 0.57) %	ns>ctls
	31 HD		2.1 (1.5) D		4.43 (SE 0.71) %	sig.>ctls & preHD	1.87 (SE 0.43) %	sig.>ctls
	29 ctls					-0.39 (SE 0.42) %		0.002 (SE -0.47)%
SD = standard deviation; YTO = predicted years to onset; D = years duration; yr = year; ctls = controls; prog. = progressors, non-prog. = non-progressors; Manual = manual delineation; sig. = significantly; ns = not sig.; diff. = different; SE = standard error; CBSI = caudate boundary shift integral; Aylward = YTO equation (Aylward et al. 1996); Langbehn = YTO equation (Langbehn et al. 2004).								

2.1.2 Whole-Brain

In manifest HD, significantly increased whole-brain atrophy rates have been consistently detected relative to controls (Table 2-2 (Henley et al. 2006;Henley et al. 2009;Tabrizi et al. 2011;Tabrizi et al. 2013;Wild et al. 2010)). Findings are less consistent in preHD with several studies failing to detect significant increases in rates compared with controls (Henley et al. 2009;Wild et al. 2010). Lack of consistency here is likely to be a result of differences between studies with respect to sample size (those that found significant rates of whole-brain atrophy in preHD had larger cohorts (Aylward et al. 2011;Majid et al. 2011b;Tabrizi et al. 2011;Tabrizi et al. 2012;Tabrizi et al. 2013)) and the characteristics of the cohorts studied (especially YTO). Over 24 months TRACK-HD found progressors to have significantly increased rates of whole-brain volume loss than non-progressors, suggesting an association with clinical progression (Tabrizi et al. 2012).

Table 2-2. Summary table of longitudinal observational MRI studies in HD reporting whole-brain volume change over time.

Study Reference	Cohort	Interval (Months)	YTO/Years Duration (D)	Method	Whole-Brain Volume Loss	Statistical Sig.
	n	Mean (SD)	Mean (SD)		Mean (SD)	
(Henley et al. 2006)	13 HD 7 ctls	5.9 (0.6) 5.7 (0.6)	3.5(2.7) D	BBSI (Freeborough and Fox 1997)	1.1 (0.88) %/yr 0.26 (0.54) %/yr	sig.>ctls
(Puri et al. 2008)	16 HD (drug) 18 HD (placebo)	12	Stage I or II *Clinical trial	SIENA (Smith et al. 2001;Smith et al. 2002)	0.75% 1.22%	sig. diff. between groups
(Henley et al. 2009)	17 preHD 27 HD 18 ctls	12 (1.2)	16.8 (6.6) YTO 4.5 (2.9) D	BBSI	0.06 (0.47) %/yr inc. 0.92 (0.67) %/yr 0.38 (0.51) %/yr	ns diff. from ctls sig.>ctls
(Wild et al. 2010)	17 preHD 21 HD 10 ctls	27.6 (1.2) 26.4 (1.2)	17.6 (7.3) YTO 4.9 (2.6) D	BBSI	0.22 (0.23) %/yr 0.88 (0.5) %/yr 0.16 (0.25) %/yr	ns>ctls sig.>ctls
PREDICT-HD (Aylward et al. 2011)	82 preHD (far) 73 preHD (mid) 56 preHD (near) 60 ctls	24.4 (1.6) 24.4 (1.4) 24.6 (1.9) 24.1 (1.1)	>15 YTO 9-15 YTO <9 YTO	BRAINS2 (Magnotta et al. 2002)	6.62 (17.53) ml 14.68 (20.76) ml 21.37 (16.24) ml 2.52 (17.39) ml	sig.>ctls sig.>ctls sig.>ctls
(Majid et al. 2011b)	35 preHD 22 ctls	12 (1.2)	5.8 (6.9) YTO (Aylward) or 14 (7) YTO (Langbehn)	SIENA	0.388 (0.425) % 0.149 (0.348) %	sig.>ctls
TRACK-HD (Tabrizi et al. 2011)	116 preHD 114 HD 115 ctls	11.5 (0.8) 11.6 (0.8) 11.6 (0.8)	10.8 YTO Stages I & II	BBSI	0.2 (0.05 to 0.34) %/yr 0.6 (0.44 to 0.76) %/yr 0.3 %/yr	sig.>ctls sig.>ctls
TRACK-HD (Tabrizi et al. 2012)	33 preHD prog. 78 preHD non-prog.	24	10.8 YTO	BBSI	1.083 (0.703) % 0.604 (0.788) %	sig. diff. between groups
IMAGE-HD (Dominguez et al. 2013)	31 preHD 31 HD 29 ctls	18	14.7(8) YTO 2.1 (1.5) D	FSL (Jenkinson et al. 2012)	0.99 (SE 0.24) % 1.81 (SE 0.31) % -0.16 (SE 0.22) %	sig.>ctls sig.>ctls & preHD
SD = standard deviation; YTO = predicted years to onset; D = years duration; yr = year; ctls = controls; prog. = progressors, non-prog. = non-progressors; sig. = significantly; ns = not sig.; diff. = different; SE = standard error; BBSI = brain boundary shift integral; Aylward = YTO equation (Aylward et al. 1996); Langbehn = YTO equation (Langbehn et al. 2004).						

2.1.3 Ventricles

Ventricular measures are complementary to whole-brain measures, since both reflect the global effects of neurodegeneration. In HD however ventricular enlargement to some degree reflects local atrophy as the caudate is adjacent to the ventricles. Mean ventricular expansion of 1.44ml in the year immediately before disease onset and 1.57ml in the year immediately after disease onset has been reported (Hobbs et al. 2010a). These findings are supported by a larger study showing ventricular expansion 0.42ml greater in preHD than controls and 1.63ml greater in HD than controls over 12 months (Tabrizi et al. 2011). In the TRACK-HD study, there was a significant difference in the rate of expansion between preHD progressors and non-progressors (Tabrizi et al. 2012), all of which suggests an increased rate of atrophy and consequent ventricular enlargement with disease progression. Although ventricular markers are appealing as measurement is facilitated by well-defined boundaries, there is large natural variability in ventricular volume between individuals. There is also some suggestion that ventricular measurements may be more affected by non-disease related changes such as dehydration, hydrocephalus and diuretic therapy (Schott et al. 2005), which could confound some subtle longitudinal changes.

2.1.4 Grey and White Matter

Voxel-based morphology (VBM) studies that analyse tissue volume change across the whole-brain, have reported significantly elevated atrophy rates in subcortical GM and selective cortical regions in HD (Hobbs et al. 2010b; Ruocco et al. 2008; Tabrizi et al. 2011; Tabrizi et al. 2012; Tabrizi et al. 2013) and preHD (Kipps et al. 2005; Ruocco et al. 2008; Tabrizi et al. 2011; Tabrizi et al. 2012; Tabrizi et al. 2013). Whole-brain GM atrophy appears more difficult to detect using ROIs, with only one of three studies managing to detect significant GM atrophy in preHD (Dominguez et al. 2013).

More recent work suggests that WM, as well as GM, changes are also important in HD with widespread elevated WM atrophy rates reported using VBM in manifest HD compared with zero (Ruocco et al. 2008) and compared with controls (Hobbs et al. 2010a; Tabrizi et al. 2011). In preHD WM atrophy rates were significantly elevated compared with controls in a large cohort (Tabrizi et al. 2011) but a smaller study failed to detect a difference (Hobbs et al. 2010b). Studies delineating the WM have detected significant longitudinal volume loss many years before disease onset relative to zero (Ciarmiello et al. 2006; Squitieri et al. 2009a) and to control rates (Aylward et al. 2011). PREDICT-HD found the frontal lobe WM to be disproportionately affected (Aylward et al. 2011). In fact, when normal age-related atrophy was taken into account (i.e. change observed in the control group), WM atrophy was greater than striatal atrophy in the preHD group, implicating this as a strong, but largely unexplored biomarker candidate. In TRACK-HD, the preHD progressors were found to have significantly higher rates of both GM and WM atrophy than non-progressors suggesting an important correlation between neuroimaging changes and clinical decline (Tabrizi et al. 2012). Table 2-3 summarizes these results.

2.1.5 Correlations with Clinical Measures

One potential limitation of structural MRI markers as outcome measures is a lack of certainty over how they relate to clinical decline in HD. Many studies have investigated associations between structural imaging and clinical measures but the relationship is often complicated by noise in both domains, a lack of sensitivity to change in either or both, and/or a temporal dissociation between the two, e.g. structural degeneration is evident over a decade prior to the onset of overt clinical signs. It has required large multi-site observational studies, such as TRACK-HD, to provide sufficient power to address these relationships. TRACK-HD reported a significant association between atrophy rates and TFC and TMS (Tabrizi et al. 2011;Tabrizi et al. 2012). More extensive and focused investigations of the relationship between MRI metrics within clinical and cognitive measures (quantitative motor, oculomotor, cognitive and neuropsychiatric) have been cross-sectional (Bechtel et al. 2010;Scahill et al. 2011). These studies have shown strong associations between MRI and disease presentation.

2.1.6 Summary: Structural MRI in HD

Large-scale structural MRI studies have consistently detected global and local atrophy in HD gene-carriers over a decade before symptom onset. Longitudinal imaging of caudate volume is one of the most promising biomarkers for future trials as measurement is reliable and reproducible and the structure is disproportionately affected by the disease. The putamen is also a strong candidate but there is a possibility that, due to poorer tissue contrast, measurement methods are under-performing in this region. This will be investigated in Section 7.3. Global measures such as the whole-brain and ventricles show slower rates, most likely because they include regions not yet recruited by the disease. These larger regions however have the advantage, over local regions, of showing the global effects of any intervention.

Table 2-3. Summary table of longitudinal observational MRI studies in HD reporting GM and WM volume change over time.

Study Reference	Cohort	Interval (Months)	YTO/Years Duration (D)	Method	GM Volume Loss	Statistical Sig.	WM Volume Loss	Statistical Sig.
	n	Mean (SD)	Mean (SD)		Mean (SD)		Mean (SD)	
(Ciarmiello et al. 2006)	10 preHD	16.2 (3.3)		Multispectral relaxometric approach (Alfano et al. 1997)		ns>0		sig.>0
	21 HD	18.4 (4.2)	7.7 D			ns>0		ns>0
(Squitieri et al. 2009a)	11 preHD			(Alfano et al. 1997)	0.03 ml/yr	ns>0	0.87 ml/yr	sig.>0
	21 HD: Stage I				0.75 ml/yr	sig.>0	0.35 ml/yr	ns>0
	Stage II	21.32 (8.47)			1.24 ml/yr	sig.>0	0.22 ml/yr inc.	ns>0
	Stage III				1.15 ml/yr	ns>0	1.04 ml/yr	ns>0
	48 ctls				0.01 ml/yr	ns>0	0.02 ml/yr	ns>0
(Squitieri et al. 2009b)	12 HD (drug)			(Alfano et al. 1997)	sig. less atrophy			ns diff. between groups
	11 HD (placebo)	24	*Clinical trial		sig. more atrophy			
PREDICT-HD (Aylward et al. 2011)	82 preHD (far)	24.4 (1.6)	>15 YTO	BRAINS2 (Magnotta et al. 2002) *cortical GM only	0.8 (12.88) ml	ns>ctls	5.15 (15.13) ml	sig.>ctls
	73 preHD (mid)	24.4 (1.4)	9-15 YTO		0.44 (14.6) ml	ns>ctls	12.71 (18.12) ml	sig.>ctls
	56 preHD (near)	24.6 (1.9)	<9 YTO		2.17 (12.8) ml	ns>ctls	17.42 (15.15) ml	sig.>ctls
	60 ctls	24.1 (1.1)			3.04 (15.37) ml		0.99 (13.51) ml inc.	
TRACK-HD (Tabrizi et al. 2012)	33 preHD prog.			VBM (SPM)	0.40 (0.45) %	sig. diff. between groups	2.62 (0.57)%	sig. diff. between groups
	78 preHD non-prog.	24	10.8 YTO		0.22 (0.35) %		1.86 (0.64)%	
IMAGE-HD (Dominguez et al. 2013)	31 preHD		14.7(8) YTO	FSL (Jenkinson et al. 2012)	1.48 (SE 0.38) %	sig.>ctls	0.44(SE 0.46)%	ns>ctls
	31 HD	18	2.1 (1.5) D		2.10 (SE 0.43) %	sig.>ctls	1.57(SE 0.46)%	sig.>ctls
	29 ctls				0.07 (SE 0.37) %		-0.50(SE 0.38)%	

SD = standard deviation; YTO = predicted years to onset; D= years duration; yr = year; ctls = controls; prog. = progressors, non-prog. = non-progressors; sig. = significantly; ns = not sig.; sig. diff. = different; SE = standard error; inc. = increase; BBSI = brain boundary shift integral; VBM = voxel-based morphometry.

2.2 Diffusion MRI

DTI (diffusion tensor imaging) is a technique that has emerged over the last 20 years as a valuable tool for investigating tissue integrity and connectivity *in vivo* (Song et al. 2003). More specifically, diffusion markers are thought to reflect the structural stability of neural tracts within the brain by detecting the extent and coherence of water diffusion. Diffusion characteristics are altered by the degenerative process and therefore DTI facilitates the detection of micro-structural neuropathology. For inherent biological reasons, it is assumed that micro-structural measures may show improved sensitivity to HD-related pathology compared with macro-structural measures, i.e. it would be expected that disruption of cellular membranes and axonal injury would precede gross morphometric changes.

During diffusion image acquisition MR is applied at multiple angles thereby measuring the amount of water diffusing at each of these angles in each image voxel. Tissues demonstrate diffusion anisotropy, which is the property of having different values when measured in different directions. These values are used to determine the diffusion tensor: a matrix characterizing the 3D diffusion pattern within each voxel. Tensors are composed of vectors quantifying water diffusion in three directions; the principle diffusion direction (PDD) and two additional directions. The relationship between the eigenvalues of these vectors reflects the characteristics of diffusion (ranging from isotropic to anisotropic). Diffusion is largely isotropic through the fibrous WM, more diffuse within the layered GM and fully anisotropic within the viscous cerebrospinal fluid (CSF).

Table 2-4. A summary of quantitative measures of diffusion MR images.

Measure	Abbreviation	Description
Fractional Anisotropy	FA	A measure of the coherence of diffusion.
Mean Diffusivity/Trace	MD	A directionally averaged measure quantifying mean diffusion.
Axial Diffusivity	AD	A measure of diffusion parallel to WM fibres.
Radial Diffusivity	RD	A measure of diffusion perpendicular to the PDD.

Quantitative measures (summarized in Table 2-4) are derived from the diffusion tensors: mean diffusivity (MD), fractional anisotropy (FA), axial diffusivity (AD) and radial diffusivity (RD). FA is a measure of the directional coherence of diffusion, i.e. the dominance of the PDD over the other directions. Within healthy, fibrous WM tracts this value is high as a result of stable water diffusion along, rather than ‘leakage’ out of, the tract. MD quantifies diffusion in all directions within each tensor. Increased MD can be interpreted as an increase in the average spacing between membrane layers and ‘leakage’ out of WM tracts, potentially due to demyelination. AD and RD are measures of diffusion parallel and perpendicular to the PDD respectively. Increased RD is thought to reflect demyelination (Song et al. 2002). Diffusion findings within GM are more difficult to interpret as GM is composed of layered, rather than fibrous, tissue (Rulseh et al. 2013). Interpretation is currently limited to the conclusion that GM diffusion changes imply a pathological

change within the micro-structure or composition of the tissue resulting in abnormal and disorganised water diffusion.

Studies employ either a ROI-approach, where metrics are averaged over that particular region, or an automated voxelwise-approach, such as Tract-Based Spatial Statistics (TBSS; <http://www.fmrib.ox.ac.uk/fsl/tbss/index.html>). With TBSS the metric is compared between groups within a WM skeleton containing only the major WM tracts.

Despite diffusion metrics often being suggested as biomarkers of disease progression based on cross-sectional findings, only four published studies in HD to-date have investigated diffusion imaging in a longitudinal setting (Table 2-5). A recent study (the PADDINGTON study; unpublished) also assesses longitudinal DTI in HD – results of which are described in Chapter 18.

2.2.1 Reported Findings

Table 2-5. A summary table of the longitudinal DTI studies in HD.

Study Reference	Cohort	Interval (Months)	YTO/Years Duration (D)	Acquisition Parameters	Regions Studied	Diffusion Metric	Results
	n	Mean (SD)	Mean (SD)	Directions (n) Thickness (mm)			(Change per Year)
(Weaver et al. 2009)	7 ctls 4 preHD 3 HD	12	13.2 YTO 2.67 D	32 2mm	TBSS skeleton	FA AD RD	ctls: ns>0 preHD & HD: decrease sig.>0 ctls: ns>0 preHD & HD: decrease sig.>0 ctls: ns>0 preHD & HD: increase ns>0
(Vandenberghe et al. 2009)	8 HD	25 (3.1)	1.5(2.1) D	3 4mm	Caudate Putamen	MD	ns>0 ns>0
(Sritharan et al. 2010)	16 ctls 17 HD	12	5.4(3.6) D	28 2.5mm	CC Caudate Putamen Thalamus	MD	ns>0 or ctls ns>0 or ctls ns>0 or ctls ns>0 or ctls
IMAGE-HD (Dominguez et al. 2013)	29 ctls 31 preHD 31 HD	18	14.7(8) YTO 2.1 (1.5) D	60 2mm	Caudate Putamen	MD FA MD FA	No sig. between-group diffs. HD decrease sig.>ctls HD increase sig.>preHD, ns>ctls No sig. between-group diffs.
SD = standard deviation; YTO = estimated years to onset; D = years duration; sig. = significantly; diffs. = differences; ctls = controls; ns = not significantly; TBSS = tract-based spatial statistics; CC = corpus callosum.							

Using TBSS significant reductions in FA and AD (from zero) were detected throughout the brain of HD participants over one year (Weaver et al. 2009). Changes were particularly notable within callosal and frontostriatal tracts. There were also some regions of (non-significant) RD increases, suggestive of demyelination. This study however was small (n = 14) and used a very limited number of directions for acquisition. In contrast, despite larger numbers, two longitudinal ROI studies failed to detect significant changes over one or two years (Sritharan et al. 2010; Vandenberghe et al. 2009). The IMAGE-HD study did find, over 18 months, significant caudate FA decrease in the manifest (but not premanifest) HD group

compared with controls (Dominguez et al. 2013). Significant MD decrease was also found in the putamen of the preHD group compared with the manifest HD group; a finding which is difficult to interpret.

2.2.2 Summary: Diffusion MRI in HD

Theoretically DTI may offer improved sensitivity compared with macro-structural volumetric analysis, however longitudinal research is at a very early-stage of development. The inconsistency in current findings is most likely influenced by the methodology used. Sample size and heterogeneity are also major limitations of these first published longitudinal studies.

2.3 Functional MRI (fMRI)

Functional imaging (Buxton 2002) uses Blood Oxygenation Level Dependent (BOLD) signal to localise regions of brain activated by a task, discover brain networks that function together to complete certain tasks, and can be used to detect functional abnormalities in individuals with HD compared with controls (Paulsen 2009). There is a suggestion that functional MRI (fMRI) may be able to pick up early neural dysfunction before morphological changes take place and hence may provide a more sensitive measure than structural or diffusion MRI.

Cross-sectional functional studies in preHD have reported lower task-related activations even when performance levels were normal (Wolf et al. 2007). This was sometimes also accompanied by enhanced cortical activation, often interpreted as neural compensation for dysfunctional circuitry elsewhere (Papoutsis et al. 2014). To-date only one fMRI study has managed to detect significant change over time in HD (Georgiou-Karistianis et al. 2013a). This study detected significant increases (interpreted as compensation) in functional activations in a preHD group compared with controls over an 18-month period during performance of a working memory task. These increases were not detected in the manifest HD group.

The lack of longitudinal fMRI studies currently published in the literature may be due to methodological limitations, either in image acquisition (difficulties ensuring reliability or maximising the signal-to-noise ratio), study design or data analysis. Test-retest reliability, multi-site and cross-culture studies are needed to increase confidence in functional change scores. Improvements in scanner stability and technical aspects of functional imaging may be necessary. It might also be that cognitive dysfunction in HD does not evolve uniformly making this unfeasible as a biomarker.

fMRI is not used in this thesis but large multi-site studies, such as TRACKON-HD (<http://hdresearch.ucl.ac.uk/current-studies/trackon-hd/>) which is currently underway, seek to address the utility of this imaging modality to track disease progression.

2.4 PET

Positron emission tomography (PET) involves the injection of a radio-labelled ligand specifically designed to bind to particular structures or substances within the brain which can be imaged to detect metabolic and neural changes. PET studies in HD typically quantify dopamine receptor binding potential or measure glucose metabolism. These are thought to reflect availability of the striatal medium spiny neurons and disturbance of glucose metabolism respectively. Although PET shows sensitivity and is likely to play a role in future trials requiring pharmacodynamic markers, the invasive nature and expense of this modality make it less suitable for large longitudinal observational studies. For that reason, PET will not be considered further in this thesis.

2.5 Magnetic Resonance Spectroscopy (MRS)

Magnetic resonance spectroscopy (MRS) is another promising MR-based biomarker in HD, although in very early stages of development. MRS is able to detect concentrations of brain metabolites and has detected reduced levels in HD (Sturrock et al. 2010). No study has yet published longitudinal results from MRS in HD.

2.6 Neuroimaging Biomarkers in Clinical Trials

Imaging biomarkers are being increasingly used to provide surrogate end-points for phase II clinical trials in HD. In these trials the aim is to ascertain safety and tolerability but also to gather initial proof-of-concept data as evidence of biological effect, i.e. that the agent may have neuroprotective properties. The first of these trials are summarised in Table 2-6. The majority of these trials reported, or are planning to output, volumetric MRI or PET read-outs.

2.7 Neuroimaging in HD: Literature Summary

As previously described, the characteristics of an ideal biomarker include reliability, reproducibility, minimal invasiveness, wide availability, low variability in the normal population, linear change with disease progression and predictable response to an intervention which modifies the disease. The imaging modalities that currently best fit this description are structural MRI and PET. MRI volumetric biomarkers have been shown to be reliable, widely reproducible and sensitive to disease progression. PET is more expensive, less widely available and may not be as sensitive to longitudinal change but has the advantage of being able to target specific molecules. Accordingly, these are the two most common imaging modalities used in the first trials in HD to utilise imaging biomarkers. Longitudinal DTI and fMRI require further exploration in large multi-site observational studies and optimisation in terms of repeatability, reliability and longitudinal signal-to-noise before their more widespread use in a clinical trial environment should be considered. However both show potential in terms of scientific interest and as biomarkers.

Table 2-6. Details of clinical trials in HD to-date utilising neuroimaging biomarkers.

Therapy	Trial Name	Reference	N with Imaging	Interval (months)	Neuroimaging Biomarker	Imaging Results	Follow-Up
Lamotrigine		(Kremer et al. 1999)	26	30	FDG-PET	No significant treatment effect was detected. Both placebo and treatment groups showed significantly decreased regional metabolism over time.	No follow-up planned.
Ethyl-EPA	TREND-HD	(Puri et al. 2002)	4	6	Registered difference MR images	Placebo was associated with progressive cerebral atrophy (evident around the ventricles) whereas the drug group showed the reverse process.	Details of follow-up below.
		(Puri et al. 2008)	34	6	Registered difference MR images	Patients treated with ethyl-EPA, showed stable or improved motor function. No significant effects in the imaging.	A larger, 6-month randomised controlled trial (n=316; no imaging) failed to find any therapeutic motor improvement (Huntington Study Group TREND-HD Investigators 2008).
Co-Enzyme Q10	CARE-HD, 2CARE, PREQUEL	(Aylward et al. 2003)	19	30	Caudate volume	No therapeutic clinical or biological effect of the drug was found.	A large (n=608) phase III trial of coenzyme Q10 is currently underway (2CARE; Huntington Study Group).
Riluzole		(Squitieri et al. 2009b)	23	24	MRI and FDG-PET	Riluzole was found to protect from glucose hypometabolism, reduce GM volume loss, increase production of neurotrophins and improve clinical scores, compared with the placebo group.	This was followed-up by a large, three-year randomized control trial (n=537, no imaging). In this larger trial no beneficial therapeutic effect was found (Landwehrmeyer et al. 2007).
Memantine	MITIGATE-HD	(Hjermind et al. 2011)	4	3-4 (18 n=1)	FDG-PET	There were no neuropsychological or metabolic changes during the 3 to 4-month treatment. One patient prolonged treatment for 18 months and showed no deterioration in either cognitive or metabolic measures, whilst those that stopped treatment after three to four months had minor progression on all cognitive domains tested, suggesting there may be a slight benefit.	A phase III trial is currently underway (Huntington Society of Canada & Huntington Study Group).

Intracerebral Grafting	MIT-HD	(Paganini et al. 2013)	26	36	FDG-PET	Striatal/cortical metabolic increases were seen over 2 years. This was slightly decreased after 4 years but still higher than preoperatively	Life-time follow-up may ultimately clarify whether transplantation permanently modifies the natural course of the disease. Follow-up currently up to 5.1 years.
Creatine	PRECREST	(Rosas et al. 2014)	64	18	FreeSurfer cortical thickness & DTI	Neuroimaging demonstrated treatment-related slowing of cortical and striatal atrophy over 6 and 18 months, although no cognitive benefits were demonstrated.	To be confirmed.
PBT2	REACH2HD	Press release from Prana	4	6	MRI	Reduced atrophy of brain tissue in areas affected in HD seen in a pilot imaging sub-study	Phase III trial ongoing (Huntington Study Group).
Citalopram		No imaging results published.	33	20 weeks	MRI (striatum)	No imaging results have been published. No benefit of short-term treatment found on cognitive functioning although there is a suggestion of improved mood (Beglinger et al. 2014).	To be confirmed.
Epigallocatechin Gallate (EGCG)	ETON-Study	NCT01357681		12	MRI (VBM)	No results published.	Estimated completion date: July 2015
PF-0254920 (Pfizer compound)		NCT01806896		28 days	fMRI: Monetary Incentive Delay task	No results published.	Estimated completion date: Nov 2015

3 Thesis Methods: The PADDINGTON Study – Work Package 2

At this stage of HD research, in order to increase and promote the usage and utility of neuroimaging biomarkers in upcoming clinical trials, direct comparisons between different imaging biomarkers are required, over short and varying time intervals – these were the goals of the PADDINGTON study.

The PADDINGTON study (*'Pharmacodynamic Approaches to Demonstration of Disease-Modification in Huntington's Disease by SEN0014196'*) was a European Seventh Framework Programme Project - full information can be found on the webpage: <http://www.paddingtonproject.eu>. This project was comprised of several work packages. Results from work package 2 (WP2) are reported in this thesis:

WP2: A multi-centre, multi-national prospective observational imaging biomarker study in early-stage HD patients to assess imaging techniques and parameters able to support efficacy studies with SEN0014196 in HD patients during Phase II and III studies.

WP2 focused on 3 Tesla (3T) diffusion and structural MRI. Its aim was to address logistical challenges and critical issues such as repeatability, short-interval scanning and stability of image data acquisition in a multi-centre setting, thus laying the groundwork for the use of advanced MR neuroimaging techniques in multi-site, multi-investigator clinical trials.

3.1 Cohort

Participants (61 early-stage HD patients and 40 age-matched controls) were recruited across four European sites (Leiden, London, Paris and Ulm). There were three study visits; at baseline, six and 15 months. All controls and 59/61 HD patients returned for the 6-month assessment. At the 15-month assessment 37/40 controls and 56/61 HD patients returned.

The majority of the control participants were spouses or normal repeat-length siblings of the HD participants. Age and gender were well-balanced between groups, by design. For the controls, each of the four study sites contributed exactly 25% of the sample. For the HD participants, contribution from each site ranged between 21% and 28%. 5/61 HD participants did not fulfil criteria for being within Stage I of the disease; four of these were Stage 2 and one was Stage 3. Full participant demographics are detailed in Table 3-1.

HD participants were required to have had a positive genetic test ($CAG \geq 36$), be able to tolerate MRI and sample donation, and have no clinically significant and relevant history that could affect the conduct of the study and evaluation of the data; as ascertained by the investigator through a detailed medical history.

The presence of major psychiatric disorder, a concomitant significant neurological disorder, significant medical illness, unwillingness to donate blood and/or concurrent participation in a clinical drug trial

disqualified entry to this study. Participants were also assessed for unsuitability for MRI e.g. claustrophobia, metal implants, cardiac pacemakers or history of significant head injury.

Table 3-1. PADDINGTON study participant demographics – adapted from Hobbs et al. (Hobbs et al. 2013).				
Characteristic	Controls (n=40)		HD Stage I (n=61)	
Age (Years)				
Mean (SD); Range	51.4 (8.4)	29.0 - 66.6	48.7 (10.8)	23.5 - 67.3
Gender				
Female n (%)	23	(57.5%)	37	(60.7%)
Male n (%)	17	(42.5%)	24	(39.3%)
Centre				
Leiden n (%)	10	(25%)	17	(28%)
London n (%)	10	(25%)	16	(26%)
Paris n (%)	10	(25%)	13	(21%)
Ulm n (%)	10	(25%)	15	(25%)
TMS				
Mean (SD); Range	1.4 (1.9)	0 - 7	20.1 (10.7)	6 - 58
TFC				
Mean (SD); Range	12.98 (0.16)	12 - 13	11.74 (1.45)	5-13
CAG				
Mean (SD); Range	NA		43.8 (3.2)	39 - 54
Disease Burden Score ^a				
Mean (SD); Range	NA		376.5 (85.2)	226.4 - 559.2
TFC Breakdown				
n (%)	NA			
TFC 11-13 (HD Stage I)			56	(91.8%)
TFC 7-10 (HD Stage 2)			4 ^b	(6.6%)
TFC 3-6 (HD Stage 3)			1 ^c	(1.6%)
^a Disease-burden formula (Penney et al. 1997): age x (CAG – 35.5); ^b 3 London site, 1 Paris site; ^c Paris site. NA= not applicable.				

3.2 Ethical Approval

All participants gave written informed consent in accordance with the Declaration of Helsinki. For the London site REC approval was given by the Central London REC 4.

3.3 Assessments

Assessments were conducted at baseline, six and 15 months during visits that lasted roughly three and a half hours. During these visits the following assessments were conducted:

- **Clinical assessment:** TMS, TFC and functional assessment sections of the UHDRS.
- **Past medical history:** Birth trauma or neonatal illness, childhood illness, adult illness, surgery, alcohol units per week, alcohol status (never abused; previous abuse; current abuse), recreational drug use, tobacco (current; ex; never), cigarettes per day, years of smoking, allergies.

- **Medication:** Name, dose, duration for each, active medical conditions.
- **Body mass index:** Calculated for all participants at each visit.
- **HD history:** Affected parent, parental onset age, onset age, first symptom, date of genetic test, analysing laboratory, small allele length, large allele length.
- **Psychiatric history:** Previous depression, previous anxiety disorder, previous obsessive compulsive disorder diagnosis, previous psychotic illness, previous suicide attempt, previous self-harm, previous suicidal ideation.
- **Neuropsychiatric Assessment:** The HADS-SIS; a composite psychiatric score from the Hospital Anxiety and Depression Scale (HADS (Zigmond and Snaith 1983)) and the Snaith Irritability Self-assessment scale (SIS (Snaith et al. 1978)) and the Columbia Suicide Severity Rating Scale.
- **Cognitive assessment:** The core cognitive battery of REGISTRY 3, including Stroop Colour Naming, Word Reading and Interference tests, Trails A and B tasks, the Hopkins Verbal Learning Test (HVLT), Verbal and Category Fluency and the Symbol Digits Modalities Test (SDMT).

3.4 Image Acquisition

Different scanners were available at each study site therefore the acquisition parameters varied slightly. 3T structural MRI data (T1- and T2-weighted) were acquired based on protocols previously standardised for multi-site use (Tabrizi et al. 2009). Diffusion acquisition parameters were carefully calibrated and tested to ensure that data collection was as consistent as possible (Muller et al. 2013).

Leiden: Philips Achieva 3T scanner. T1-weighted magnetisation-prepared rapid gradient echo (MP-RAGE) scans were acquired with the following parameters; TR = 7.7ms, TE = 3.5ms, flip angle = 8°, field of view (FOV) = 24cm, matrix size = 224x224, yielding 164 sagittal slices to cover the entire brain with a slice thickness of 1.0 mm with no inter-slice gap. Diffusion-weighted MR images for DTI were acquired using an echo planar imaging (EPI) protocol with the following parameters; 55 axial slices of 2 mm thickness, with no inter-slice gaps, acquisition matrix = 112 x 112, in-plane resolution of 2 mm², resulting in isotropic voxels (TR = 8062 ms, TE = 56 ms). Diffusion data were acquired in 42 different encoding directions with b = 1000 s/mm², along with one b = 0 image.

London: Siemens Tim Trio 3T scanner. T1-weighted MP-RAGE scans were acquired with the following parameters; TR = 2200ms, TE = 2.2ms, flip angle = 10°, FOV = 28cm, matrix size = 256x256, yielding 208 sagittal slices with a slice thickness of 1.0 mm with no inter-slice gap. DTI data were acquired using an EPI sequence with the following parameters; 65 axial slices of 2 mm thickness, with no inter-slice gaps, acquisition matrix = 96 x 128, in-plane resolution of 2 mm², resulting in isotropic voxels (TR = 7600 ms, TE = 84 ms). Diffusion data were acquired in 42 different encoding directions with b = 1000 s/mm², along with 7 b = 0 images.

Paris: Siemens Verio 3T scanner. T1-weighted MP-RAGE scans were acquired with the following parameters; TR = 2200ms, TE = 2.2ms, flip angle = 10°, FOV = 28cm, matrix size = 256x256, yielding 208 sagittal slices with a slice thickness of 1.0 mm with no inter-slice gap. DTI data were acquired using an EPI sequence with the following parameters; 75 axial slices of 2 mm thickness, with no inter-slice gaps, acquisition matrix = 128 x 128, in-plane resolution of 2 mm², resulting in isotropic voxels (TR = 13100 ms, TE = 86 ms). Diffusion data were acquired in 42 different encoding directions with b = 1000 s/mm², along with 7 b = 0 images.

Ulm: Siemens Allegra 3T scanner. T1-weighted 3D scans were acquired with the following parameters; TR = 2200ms, TE = 2.81ms, flip angle = 9°, FOV = 28cm, matrix size = 256 x 256, giving 208 sagittal slices with a slice thickness of 1.1 mm with no gap. Acquisition time was 9 minutes. DWI was performed with an EPI sequence, each data volume consisted of 52 axial slices of 2.2 mm thickness, with no inter-slice gaps, acquisition matrix = 96 x 128, in-plane resolution of 2.2 mm², resulting in isotropic voxels (TR = 7600 ms, TE = 85 ms). Diffusion data were acquired in 47 different encoding directions with b = 1000 s/mm², along with three b = 0 images.

4 Thesis Methods: Image Analysis

This chapter will outline the image analysis methods used in this thesis. Section 4.1 describes image registration techniques which are applicable to both volumetric and diffusion analysis and is followed, in Sections 4.2 and 4.3 respectively, by a discussion of specific volumetric and diffusion MRI analysis methods.

4.1 Image Registration

For several methods used in this thesis, registration (alignment) of inter-subject scans or intra- subject serial scans to a common space is required. All registrations involve two steps:

1. Transformation: co-ordinate and voxel size information is used to calculate the optimal parameters needed for the registration.
2. Interpolation/resampling: application of the transform to the scan.

In all cases of registration it is necessary to: firstly choose the target onto which the scans are to be registered; then specify a transformation model with which to do this; stipulate the similarity or error measure that will be used to assess the ‘goodness-of-fit’; choose an interpolation strategy with which to resample the scan data; and finally decide on exit criteria.

4.1.1 Target of Registration

The original positioning of a brain in the scan FOV is called its native-space. In order to facilitate the use of consistent anatomical landmarks or group comparisons, image analysis may be conducted in a standard-space. This requires registration of all scans either to a freely available atlas or to a study-specific template. Alternatively, with longitudinal data-sets, serial scans can be registered to the baseline scan or to a common mid-way space. These instances are outlined below:

Atlases

The most commonly used atlases in HD are: 1) the Talairach atlas (Talairach et al. 1988) - created from a single participant’s (smaller than average) thick post-mortem brain slices; 2) the MNI305 (Montreal Neurological Institute) template (Evans et al. 1993) - created from an average of 305 brain scans from healthy, young, right-handed participants (239 males, 66 females, age 23.4 +/- 4.1 years). It is therefore questionable whether, in certain circumstances, these atlases are optimal for use with older or disease populations, including HD.

A Study-Specific Template (DARTEL)

This issue can be avoided by using a DARTEL (Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra) template (Ashburner 2007). From the SPM toolbox, DARTEL is a registration algorithm for creating a study-specific template i.e. an average of all the study’s scans. Registration to this average target reduces the risk of bias in a disease-control group comparison and decreases registration

errors, as smaller deformations are needed to match with an average-shaped template than a standard template. Details of DARTEL registration are given below under the heading of '*Non-Linear Registration*'.

Registration of Serial Scans

Registration of serial scans is typically done by registering the follow-up scan(s) to the baseline. Recently it has been proposed that this asymmetry in the registration process can introduce bias due to a lack of inverse consistency (Reuter and Fischl 2011). Here inverse consistency means that one expects to obtain the inverse transform when registering B to A (follow-up image to baseline) as opposed to A to B (baseline to follow-up image). It is important to treat all time-points identically and ensure that they undergo the same degree of resampling as each resample degrades the data. Several image analysis methods are beginning to adapt their processing pipelines to avoid registration asymmetry bias. These include:

- FreeSurfer software which offers a specialised longitudinal pipeline in which scans are registered to, and analysed in, a common half-way-space (Reuter et al. 2012).
- The PADDINGTON study diffusion pipeline which was designed to conduct longitudinal analyses in a half-way-space (more details are included in Section 4.3.3).

4.1.2 Transformation Models

Linear Registration

Rigid Registration

Rigid registration is the simplest form of registration. It uses six degrees of freedom (three rotations, three translations) to move and align scans into a common space. During this process the size and structure of all scans is maintained.

Affine Registration

Affine registration is more complex than rigid registration as it additionally resizes and shears images to match the size and position of the target template. This requires twelve degrees of freedom (three translations, three rotations, three zooms, three shears). During this registration all voxels in the image are treated identically, therefore although there may be an overall scaling up of voxel sizes this does not remove regionally-specific atrophy. Affine registration is used during processing of the boundary shift integral (BSI; details in Section 4.2.11).

Non-Linear Registration

Whilst rigid and affine registrations apply one transformation to the whole image (i.e. every single voxel within the image is treated the same whatever rotations, zooms etc. are required for a global match) non-linear registration treats each voxel differently in order to match scans as closely as possible. Two non-linear registrations used in this thesis are fluid and DARTEL (diffeomorphic) registrations.

Fluid Registration

In a fluid registration the warps are based on the physical model of a compressible viscous fluid; the method used in this thesis is based on a model published by Christensen et al. (Christensen et al. 1996) - details of this pipeline are included in Section 4.2.8. To allow each voxel to compress or expand as necessary fluid registration applies large numbers of degrees of freedom. Each fluid element experiences a force in the direction of increasing image similarity but is constrained by Navier-Stokes fluid equations that describe the motion of a compressible viscous fluid. This involves a combination of: local transformations which allow, for example, one area to shrink whilst another expands; Jacobian constraints to conserve topology (so the brain image doesn't break in two); prior knowledge of the likely extent of deformations in the form of Bayesian (probability) constraints; and smooth deformations to maintain topology.

The fluid registration generates a detailed deformation field for each subject. The amount of voxel-level expansion or contraction is extracted from each deformation field by computing the determinant of the Jacobian at each voxel, i.e. the determinant of the gradient of the deformation field. Voxel compression maps (VCMs) provide a visual representation of fluid registrations by representing these Jacobian determinants in colour; contraction in green-blue and expansion as yellow-red. Figure 4-1 shows examples of a good fluid registration (left) and one showing extreme geometric distortion between the serial scans resulting in poor registration (right). Plausible biological change will manifest on fluid registrations as selective regional atrophy, e.g. contraction of brain tissue, particularly the striatal regions, and expansion of the ventricular CSF. Distortions are reflected in fluids that show systematic shifts; the example below (right) shows voxel expansion across the inferior brain tissue, voxel compression across the superior brain tissue and a shift, manifested as severe contraction, of the brainstem. This is not biologically plausible for this subject.

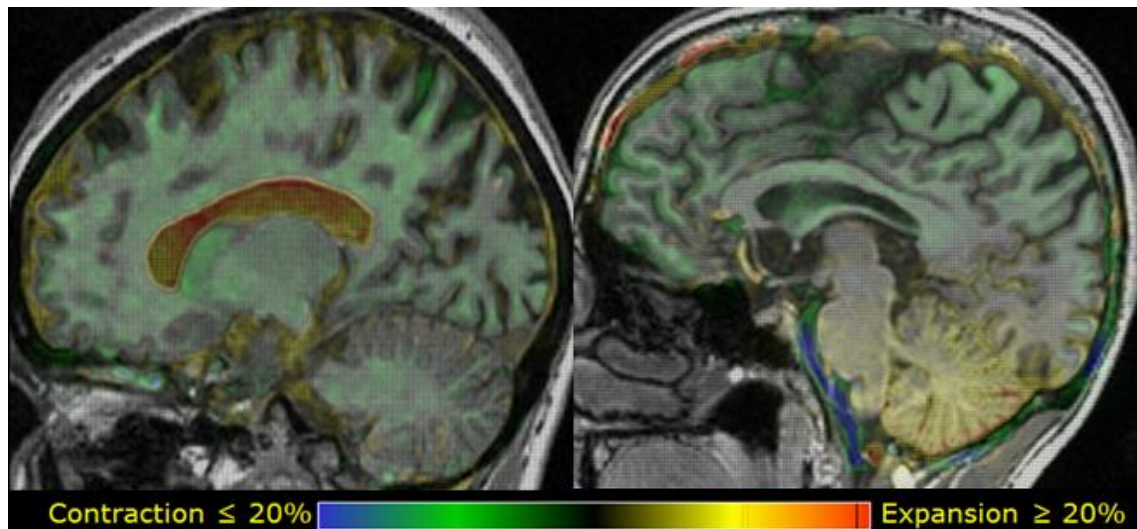


Figure 4-1. (Left) An example of a good quality VCM from one participant following fluid registration. The GM and WM can be seen to be contracting (green-blue) and the ventricular and CSF space expanding (yellow-red). (Right) A poor quality VCM showing significant distortion particularly of the cerebellar and brainstem regions.

DARTEL

DARTEL is a diffeomorphic registration algorithm for creating a study-specific template (Ashburner 2007). This method is used in SPM8 VBM analysis (details in Section 4.2.9). According to Ashburner a diffeomorphism is ‘a globally one-to-one (objective) smooth and continuous mapping with derivatives that are invertible’. In other words, this algorithm constructs very smooth, large deformation fields that can be easily inverted (reversed), whilst preserving topology (avoiding image ‘tearing’).

4.1.3 Resampling and Interpolation

Resampling, or reslicing, describes the application of the transformation parameters to the scan. This step must be conducted after each iterative transformation, in order to evaluate the goodness-of-fit, and after the final iteration to create the registered image. Resampling degrades the image quality and therefore, if applicable and possible, it is recommended to combine several transforms before performing this step.

Resampling often involves the formation of new voxels in-between a discrete set of known voxels from the original, pre-registration image; for each new voxel in the transformed image the intensity must be calculated via a process called interpolation. There are multiple interpolation methods available. These methods involve inverse transformation of the unknown data point back onto the pre-registration image and sampling of the surrounding voxels on the original image to determine the unknown voxel intensity. The interpolation methods used in this thesis include:

1. **Nearest neighbour** – this is the simplest interpolation method, taking the value of the closest voxel in the original image for the resampled voxel and not considering other neighbouring voxels. This method is used for interpolation of binary masks but is generally too simplistic for image registration.

2. **Tri-linear** – interpolated voxels are computed from the neighbouring voxels by linearly weighting these depending on distance; this is the default interpolation option in SPM8 (details in Section 4.2.9).
3. **B-Spline and Sinc** – B-spline and sinc are higher-order interpolation methods for smoothly sampling the surrounding voxel intensities using a Gaussian kernel. In SPM8 B-spline is a recommended option over the default tri-linear interpolation, although at higher computational cost. Sinc interpolation is optimal for fluid registrations (Thacker et al. 1999) and a very similar method is used for BSI affine registrations (chirp-Z interpolation with the AIR (Automated Image Registration) toolkit (Woods et al. 1998a; Woods et al. 1998b)).
4. Alternatively, a mixture of methods can be applied to reduce computational costs. For example, the fluid registrations conducted for this thesis are set to compute 300 iterations; the first 250 apply tri-linear interpolation and the final 50 are sinc interpolated.

4.1.4 Goodness-of-Fit

Similarity- or cost-functions drive the transformation models, quantifying how well aligned the intensities of the scans are following each iteration. Examples of these functions include: standard deviation (SD) of ratio image, least squares regression, correlation coefficients, sums of squared differences and mutual information. For a review see Crum et al. (Crum et al. 2004).

4.1.5 Exit Criteria

Exit criteria stipulate the requirements that must be met for a registration to be considered complete. This is typically based on convergence criteria derived from cost functions and/or a defined limit to the number of iterations. For example, as previously mentioned, the fluid registrations conducted for this thesis are limited to compute 300 iterations. There is also a convergence criterion of mean body force 5×10^{-8} . This number of iterations has been empirically shown, in HD, to be sufficient and of negligible difference compared to registrations run with more iterations. The criterion is rarely reached before the 300 iteration limit.

4.2 Structural MR Image Analysis

Structural T1-weighted scans provide volumetric information about the brain. The whole-brain or selected regions can be analysed to assess volume or thickness. When there are serial scans change in these measures over time can be quantified. There are different ways of preparing and analysing T1-weighted data. The methods used in this thesis are outlined below.

4.2.1 Pre-Processing

Bias Correction

MR image analysis techniques rely on consistent voxel intensities across an image and between scans. Raw T1-weighted scans however often suffer from intensity non-uniformity (also referred to as bias, intensity

inhomogeneity, or shading artefact), i.e. images show smooth spatially varying signal intensity – an example of which is shown in Figure 4-2. This non-uniformity can be more pronounced at higher field strengths due to the shorter radio-frequency wavelength, or when using multichannel receiver coils. Bias correction methods aim to remove these intensity fluctuations. All T1-weighted scans analysed in this thesis were bias-corrected using the non-parametric non-uniform intensity normalization (N3) method of Sled et al. (Sled et al. 1998), with optimised parameters for 3T data as outlined in Boyes et al. (Boyes et al. 2008).

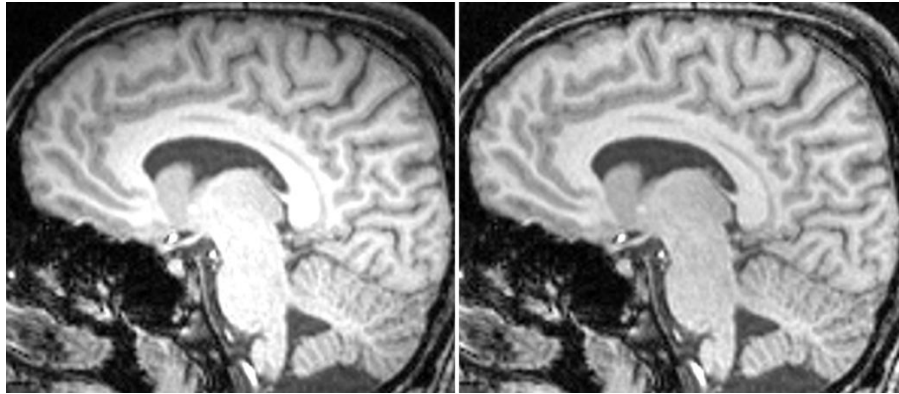


Figure 4-2. Voxel intensity fluctuations before (left) and after (right) N3 bias correction.

Bias correction between serial scans is known as differential bias correction. Here both images are used to identify any residual low-frequency variation in the ratio of the image intensities and corrected for this (Lewis and Fox 2004). This is a feature of the fluid registration and BBSI processing pipelines used in this thesis (details in Sections 4.2.8 and 4.2.11 respectively).

Quality Control (QC)

Visual QC was performed on all T1-weighted scans. The most commonly identified artefacts were due to motion, e.g. ringing, ghosting or blurring; an example of which is shown in Figure 4-3a. Intensity homogeneity and tissue contrast were also examined. Poor quality scans were rejected at the earliest stage and a rescan requested. Scans that passed initial QC but subsequently failed the serial brain-brain affine registration, most commonly due to inconsistent head positioning in the FOV (e.g. Figure 4-3b) and consequent geometric distortion, were also rescanned or the brain BSI, described in Section 4.2.11, was not used.

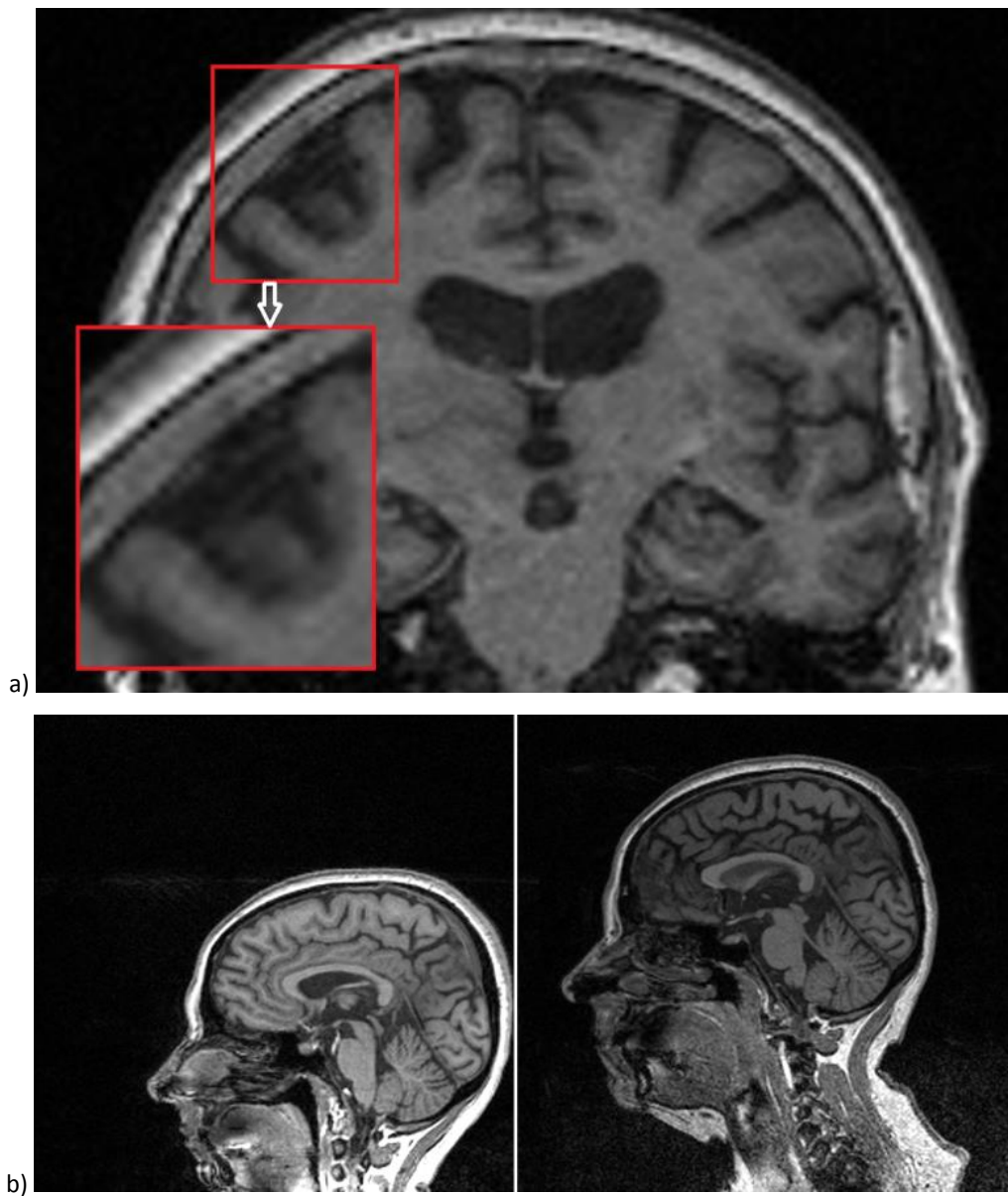


Figure 4-3. a) An example of motion artefacts, in this case ringing and blurring; b) one participant's scans at baseline and follow-up, with inconsistent head positioning in the FOV.

4.2.2 Manual Delineation

Manual delineations were conducted in MIDAS (Medical Information Display and Analysis System) software (Freeborough et al. 1997). These methods are not fully manual as analyses of each structure are initialised using pre-defined intensity constraints related to the mean brain intensity of each scan. This creates an initial outline of the structure which is subsequently refined by expert raters following detailed protocols validated for use in atrophied and healthy brains.

Details of these protocols for each ROI are outlined below and examples are shown in Figure 4-4. Manual delineations, excluding the native-space whole-brain analysis, were conducted in MNI305 standard-space. This enabled consistent application of landmark-derived cut-offs included in the segmentation protocols. Change within several regions was quantified using the BSI, details of which can be found in Section 4.2.11.

One caveat to mention with manual delineations is that although theoretically blinded to disease group, in neurodegenerative disease where atrophy is visible, the analyst conducting the delineations may be aware of the participant's disease status thereby potentially introducing bias.

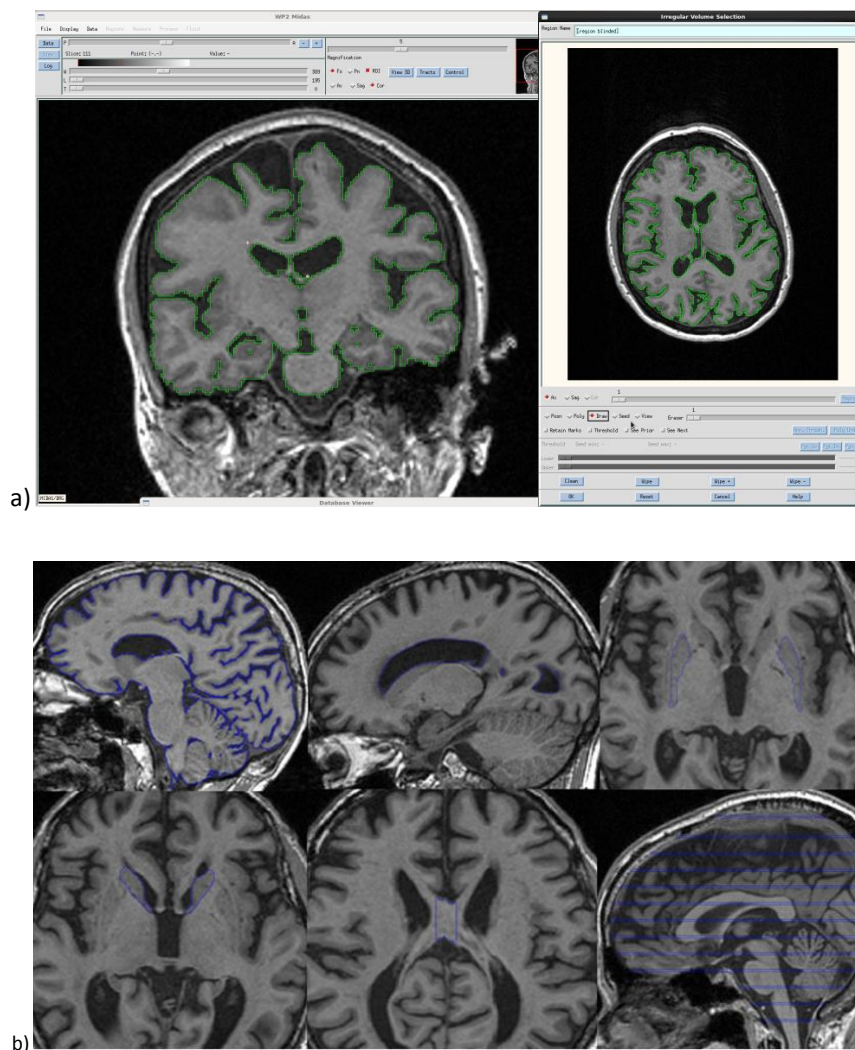


Figure 4-4. a) A screen shot of a whole-brain segmentation in MIDAS software. b) Examples of regional segmentations of (top left to right) whole-brain, lateral ventricles, putamen, (bottom left to right) caudate, corpus callosum and total intracranial volume.

Whole-brain

In native-space the brain was delineated at baseline and follow-up using a morphological segmentor which involved the application of interactive thresholds and a series of erosions and dilations to separate brain tissue from scalp and CSF. This was followed by manual editing where appropriate. The BSI, quantifying change over time, was computed using both baseline and follow-up delineations; this is reported as the brain BSI (BBSI). Each whole-brain delineation takes approximately one hour.

Lateral Ventricles

Lateral ventricle segmentations were conducted using an upper intensity threshold of 60% of the mean brain intensity and manual edits to retain the temporal horn of the lateral ventricles but not the third or fourth ventricles (Scahill et al. 2003). Delineations were conducted at baseline and follow-up after which

the ventricular BSI was computed; this is reported as the ventricular BSI (VBSI). Each delineation of the lateral ventricles takes approximately twenty minutes.

Caudate

Caudate segmentation included the head and body of the caudate, with the medial border defined by the lateral ventricle, and the lateral border defined by the internal capsule (Hobbs et al. 2009). The thresholds included voxel intensities between 62% and 111% of mean brain intensity. The caudate was delineated on the baseline scan only and the BSI was used to quantify change; this is reported as the caudate BSI (CBSI). Each caudate delineation takes approximately one hour.

Corpus Callosum (CC)

Segmentation of the corpus callosum (CC) was performed in the sagittal plane and extended four slices either side of the mid-sagittal slice for each participant; with voxel intensity thresholds set at 100% and 150% of mean brain intensity. Each delineation therefore included a total of nine image slices.

Segmentations were performed at serial time-points and an indirect measure of volume change over time was calculated by subtraction of baseline from follow-up. Each delineation of the corpus callosum takes approximately twenty minutes.

Putamen

A manual delineation protocol was developed for the putamen as part of this thesis (Section 7.2; Appendix Section 23.4). Thresholds were set at 90% and 112% of mean brain intensity, the anterior and lateral borders were defined by the WM of the internal and external capsules respectively, whilst voxel intensity differentiated the putamen from the globus pallidus along the medial border. Each putamen delineation takes approximately one hour.

Total Intracranial Volume (TIV)

Total intracranial volume (TIV) is the volume within the cranium (skull) which includes the brain, meninges and CSF. TIV can be used to adjust for inter-subject head-size variation. TIV has been shown to positively associate with regional brain volumes (Barnes et al. 2010) therefore volumes can be reported as proportions of TIV, or group comparisons can be adjusted for TIV, to remove this variability.

Image analysis of TIV involved outlining the dura, facilitated by intensity thresholds (lower threshold at 30% of mean brain intensity and no upper threshold (set to maximum)), on every tenth slice from the most inferior point of the cerebellum up to the most superior slice containing cortex. Where the intensity thresholds failed to pick up the dura this was manually delineated. The software then interpolated the volume between the analysed slices. Each TIV delineation takes approximately twenty minutes.

4.2.3 Automated Analysis

Manual delineation of brain structures, as described above, is widely accepted to be as close as possible to a 'gold standard' methodology. For practical reasons however many groups have moved to using automated methods; the most widely published of which are described below and examples of which are shown in Figure 4-5.

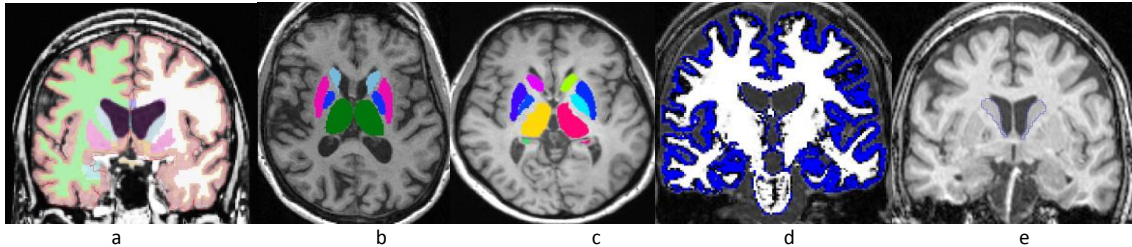


Figure 4-5. a) FreeSurfer volumetric segmentations and cortical delineation displayed with Tkmedit, b) FIRST and c) BRAINS3 subcortical segmentations and d) SPM GM and WM segmentations, all displayed in FSLView and e) a STEPS caudate segmentation displayed in MIDAS software.

4.2.4 FreeSurfer

FreeSurfer is freely available software (<http://surfer.nmr.mgh.harvard.edu/>) which provides an automated pipeline to output cortical (thickness, curvature and volume) metrics (Fischl and Dale 2000) and subcortical volumes (Fischl et al. 2002). Firstly, using a hybrid watershed/surface deformation procedure, raw T1-weighted scans undergo a skull strip which removes the skull and the majority of non-brain tissue (Segonne et al. 2004). From the skull stripped image an automated Talairach transformation (12 degrees of freedom) is calculated but the image is not resampled. This step helps to localise structures in later processes. Intensity normalization stabilises the intensity across the image (Sled et al. 1998). Once this pre-processing is completed segmentations are performed based on probabilistic information automatically estimated from a manually labelled training set which assigns one of 40 labels to each image voxel.

Cortical Thickness

Both intensity and continuity information from the segmentations and deformation procedures are used to produce representations of the cortical borders. Cortical thickness is calculated as the closest distance from the GM/WM boundary to the GM/CSF boundary at each vertex on the tessellated surface (Fischl & Dale 2000). FreeSurfer is currently the most commonly used software for cortical thickness analysis in HD.

As mentioned in Section 4.1.1, FreeSurfer offers a specialised longitudinal pipeline which aims to avoid asymmetrical registration bias (Reuter et al. 2012; Reuter & Fischl 2011). It accomplishes this by generating a template in a common space between the serial scans. All time-points are then registered to this space and processing is initialised with common information from the template.

4.2.5 FIRST

FIRST (FMRIB's Integrated Registration and Segmentation Tool (Patenaude et al. 2011)) is a model-based segmentation tool freely available from the FMRIB Software Library (FSL; <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FIRST>). It is described as a Bayesian Appearance Model. As such this method models both shape and intensity and relates these to each other within a Bayesian framework. The models were trained with 336 T1-weighted scans manually labelled with 15 subcortical structures. For the shape model the manual labels were parameterized as surface meshes and modelled as a point distribution map. Shape could then be expressed as a mean with modes of variation, i.e. the most likely variations of this shape across a training set. The intensity model maps the intensity distribution as a multivariate Gaussian, parameterized by its mean and eigenvectors (modes of variation). FIRST searches through linear combinations of shape modes of variation, based on these learned models, for the most probable shape instance given the observed intensities in a T1-weighted image. Fitting shapes to new images is done by deformable surface registration and minimising the squared difference between predicted intensities, given a shape deformation and the observed image intensities.

4.2.6 BRAINS

BRAINS (Brain Research: Analysis of Images, Networks and Systems) software (Magnotta et al. 2002) can be used to segment several subcortical structures. Images first undergo spatial normalization during which they are resampled to 1mm^3 with the inter-hemispheric fissure aligned vertically with the axial and coronal views and the line between the anterior commissure and posterior commissure aligned horizontally in the sagittal view. Next the Talairach grid is fitted and warped onto the brain. The AIR toolkit is used to align T2- and PD-weighted data to the resampled T1-weighted image. These other images are also resampled to 1mm^3 voxels. Once this pre-processing is completed the tissues are classified: a tissue classification module is applied to the three multispectral data sets and generates pure plugs for each tissue type, based on the minimum variance assumption. Plugs representing blood are also manually chosen by tracing on the image. These segmentations are then used as input for intensity normalization. An 8-bit image is generated composed of the tissue segmentations, which is then coded such that a signal intensity of 10 represents pure CSF, 130 GM, 250 WM and volumes in between represent partial volume composition of more than one tissue type, e.g. 47% GM, 53% CSF. The partial volumes are converted to discrete values based on the most likely tissue type. Neural net and atlas-based structure identification methods are used for regional segmentation: the neural net was trained based on human rater definition of the brain, caudate, putamen, cerebellum, thalamus, globus pallidus and hippocampus.

4.2.7 STEPS

STEPS stands for Similarity and Truth Estimation for Propagated Segmentations (Cardoso et al. 2013). This is a local ranking strategy for template selection based on locally normalised cross-correlation. It is an extension of the classical STAPLE algorithm by Warfield et al. (Warfield et al. 2004) which uses local

intensity features to select the best labels to fuse. A library of ‘gold standard’ manual segmentations is referenced to identify the most closely matched segmentations to the voxel intensities in the T1-weighted image. For example, if one starts from a set of 15 template images registered to the image under study, one can then calculate how well each one of the template images correlates locally with the image under study and then take only the top five templates on a voxel by voxel basis. These are then fused to form an optimal structural outline.

4.2.8 Fluid Registration

Fluid registrations based on the compressible viscous fluid model of Christensen et al. (Christensen et al. 1996) are used in this thesis for longitudinal VBM and quantification of change in GM and WM (details of these applications are included in Sections 4.2.9 and 4.2.10). These fluid registrations were run in MIDAS software between N3 bias-corrected, affine registered serial scans. These scans were additionally differentially bias-corrected to remove intensity inhomogeneity between serial scans. Scan pairs are cropped using subject-specific masks to exclude non-brain regions (e.g. neck and eyes), while including the ventricular CSF, GM, WM and a layer of brain-surface CSF. These subject-specific masks were generated by adding MIDAS-derived brain masks to MIDAS-derived ventricular masks and binarising. The resultant mask was dilated by three voxels to create ‘mask A’ and by a further two voxels to create ‘mask B’. Mask B was smoothed with a 2mm full width at half maximum (FWHM) kernel to generate an approximation to a smooth signal drop-off (mask C). The inner portion of mask C was replaced with mask A, i.e. to enforce the preservation of the original signal intensity values in the neighbourhood of eight voxels around the original brain mask, followed by a smooth drop-off around that – resulting in mask D. The scan was then multiplied by mask D to crop the brain; this was repeated for baseline and follow-up scans for each participant. The fluid registration warps each individual's repeat image to match the corresponding baseline image based on a physical model of a compressible viscous fluid (Christensen et al. 1996). Fluid registrations were run until a stopping criterion of 300 iterations was reached; the first 250 applied tri-linear interpolation and the final 50 were sinc interpolated. Previous internal validation studies have shown that this is suitable for multi-site HD data with a 12-month interval and of negligible difference compared to registrations run with more iterations. This was ascertained by inspecting the overlay of the fluid-resliced repeat scan on the baseline; good registrations created a perfect match.

4.2.9 Statistical Parametric Mapping (SPM)

Statistical Parametric Mapping (SPM; <http://www.fil.ion.ucl.ac.uk/spm/>) software can be used for tissue segmentation. For this project SPM8 was run on a Matlab R2012b platform (Mathworks, USA).

Unified and New Segmentation

The default segmentation algorithm in SPM8 is called Unified Segmentation (Ashburner and Friston 2005). This method combines tissue classification, registration and bias correction into one processing step. Tissue

segmentation involves modelling the intensity distributions by a mixture of Gaussians and using tissue probability maps (TPMs) to weigh the classification. To do this scans are iteratively warped into a standard-space which contains the TPM. TPMs have priors of where to expect certain tissue types. The New Segment toolbox is an extension of the default Unified Segmentation algorithm (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/SPM8_Release_Notes.pdf). This newer version treats the mixing proportions differently, uses an improved registration model, has the ability to process multi-spectral data and contains an extended set of TPMs which allows for a different treatment of voxels outside the brain.

Voxel-Based Morphometry (VBM)

VBM is a mass univariate analyses technique, run in SPM software, that statistically compares voxel-wise volume differences or change with no *a priori* assumptions regarding ROIs (Ashburner and Friston 2000). Statistical parametric maps (SPMs) are produced which can be used for distinguishing between-group differences, modelling change within a group or differences in change between groups, or identifying structural correlates of clinical measures.

VBM processing involved: (Unified) tissue segmentation; diffeomorphic DARTEL registration of all scans to a study-specific (DARTEL space) template (Ashburner 2007); during registration the intensity of the segmentations were modulated to maintain the individual's volumetric data which would otherwise be removed by the non-linear registration. More specifically, the intensities were rescaled depending on the amount of expansion/contraction required within each voxel to register to DARTEL space. The image was then smoothed with a variable FWHM Gaussian kernel; the aim being to reduce noise and produce data with a more normal distribution by taking a weighted average of surrounding intensities. Kernel selection should be based on the fact that analysis is most sensitive to effects that match the shape and size of the kernel (Match Filter Theorem) and smaller kernels allow for more localised effects to be detected. For this study a 4mm kernel was used to allow for adequate localisation of effects. For cross-sectional analyses voxel-wise statistical analysis was conducted on this smoothed, modulated, normalised data.

Longitudinal analyses required additional steps. Within-subject fluid registrations (details in Section 4.2.8) were computed for each participant to quantify structural change between time-points (Freeborough and Fox 1998) and outputted as VCMs. Logarithms of the VCM determinants were calculated to symmetrise the range of values around zero (i.e. no change); values below zero represent compression and values above zero represent expansion. These logged Jacobian determinants were reoriented to match the SPM8-format baseline scans and masked by the SPM GM and WM tissue segmentations. These masked regions were then warped onto the DARTEL template and smoothed using a 4mm FWHM kernel. Statistical analyses were conducted on every voxel of these smoothed, normalised, masked logged determinants resulting in SPMs. Results required adjustment for multiple comparisons – see Section 5.4.

4.2.10 GM and WM Atrophy

In the PADDINGTON study volume change was computed within the grey and WM SPM segmentations using a within-subject fluid registration approach (Christensen et al. 1996;Freeborough & Fox 1998). The processing pipeline is described below:

1. **Tissue classification** of the GM and WM was conducted on bias-corrected, native-space scans using SPM8's Unified Segmentation tool (described above). Brain-masks were generated from MIDAS-derived whole-brain and ventricle segmentations (details in Section 4.2.2) and applied to remove any remaining non-brain tissue from the GM segmentations. Once processed all segmentations were visually assessed for quality. Where processing had failed scans were reoriented to a more central and upright position within the FOV and rerun.
2. **Fluid registrations** were run in MIDAS software according to the pipeline previously detailed in Section 4.2.8.
3. **Quantification of change** within the grey- and WM was computed by convolving the fluid-derived (follow-up to baseline) VCMs for each individual with their respective segmentations in baseline native-space, giving an estimate of volume change within each tissue class.

4.2.11 The Boundary Shift Integral (BSI)

The classic-BSI is a semi-automated technique for quantifying boundary shift, e.g. atrophy or growth (Freeborough & Fox 1997). This technique was originally applied and validated in Alzheimer's Disease for the whole-brain (BBSI (Freeborough & Fox 1998)) and lateral ventricles (VBSI (Freeborough & Fox 1997)). It has more recently been extended for HD to assess caudate atrophy (CBSI (Hobbs et al. 2009)) and applied for all these regions in a large multi-site HD cohort (Tabrizi et al. 2011;Tabrizi et al. 2012;Tabrizi et al. 2013).

All scans are N3 bias-corrected (Boyce et al. 2008;Sled et al. 1998). For all BSI computations delineation of the whole-brain region is required on both baseline and follow-up scans. The VBSI pipeline additionally requires lateral ventricle delineations on both baseline and follow-up scans; the CBSI requires baseline segmentations only. These ventricle and caudate segmentations are conducted on scans rigidly registered to MNI305 standard-space.

The follow-up scan is affine registered to the baseline scan within the delineated whole-brain region, using chirp-Z (a variant of sinc) interpolation with the AIR toolkit (Woods et al. 1998a;Woods et al. 1998b). For the BBSI this is done in baseline native-space whilst for the VBSI and CBSI this is in MNI305 standard-space. Follow-up ROI delineations are resliced onto the baseline image using trilinear interpolation. In the BBSI pipeline this is followed by differential bias correction. This is not necessary for the CBSI or VBSI as long-range fluctuations are less important for local BSI computations. For the CBSI and VBSI affine registration is followed by an additional local rigid registration facilitated by the delineated baseline ROI (dilated by two

voxels) and the follow-up scan whole-brain region. For the caudate registrations this is done separately for left and right structures. These are not symmetric registrations, as described in Section 4.1.1, as the follow-up scans are registered to the baseline. Resampling is however conducted using chirp-Z (sinc) interpolation which has been shown to eliminate, or at least greatly reduce, the bias that arises from asymmetric registration (Leung et al. 2012).

The area over which the BSI is computed must cover all voxels in which the regional boundary is located/shifts, i.e. include all possible intensity transitions that would be associated with GM-CSF, WM-CSF and GM-WM substitutions. This region is created by an XOR (exclusive or) operation as follows:

- a) An intersection region, common to both baseline and follow-up segmentations, is eroded by one voxel
- b) A union region, including voxels within segmentations from baseline, follow-up or both, is dilated by one voxel
- c) Subsequently the BSI region is created from the XOR of the dilated union and eroded intersect regions, i.e. the region created from a) is subtracted from that created in b).

With only the baseline caudate segmentations this process is slightly altered for the CBSI; the left and right baseline segmentations are simply dilated and eroded by one voxel to form the CBSI computation region.

The two scan intensities are normalized by dividing voxel intensities by the mean brain intensity. The BSI is calculated using user-defined intensity windows that are the same for all scans. Each voxel in the boundary is analysed to see if it has transitioned in this way and significant intensity changes are quantified as BSI.

KN-BSI

The KN-BSI (k-means normalised-BSI (Leung et al. 2010)) offers a more robust brain atrophy measurement than the classic-BSI for longitudinal multi-site studies, by addressing differences in tissue contrast over time and between scanners. This is achieved by performing scan-pair specific intensity normalisation. This method has been shown to detect higher rates of atrophy with lower SDs than the classic method (Leung et al. 2010).

The processing pipeline is slightly different for the BBSI and VBSI compared with the CBSI. For the BBSI and VBSI the mean intensity of the GM, WM and CSF is calculated by dilating the brain region by three voxels so that CSF is included, followed by k-means clustering into three clusters (one for each tissue type). A linear regression between the mean intensities in the two scans is carried out and the results are applied to the scans to normalise the intensities. The BSI is calculated in the same way as the classic BSI but using intensity windows that were calculated specifically for the scan pair using the k-means results:

$$I_{CSF\ mean} + I_{CSF\ SD} \text{ to } I_{GM\ mean} - I_{GM\ SD}$$

*where $I_{CSF\ mean}$, $I_{CSF\ SD}$, $I_{GM\ mean}$ and $I_{GM\ SD}$ are the mean and SD of CSF intensity, and the mean and SD of GM intensity.

The linear regression may however introduce an asymmetric bias therefore the baseline and follow-up images are swapped over so both the forward and backward BSIs are calculated and averaged:

$$KN-BSI = (forward\ BSI + backward\ BSI)/2$$

*forward BSI = follow-up to baseline; backward BSI = baseline to follow-up.

The CBSI pipeline is slightly different as it is optimised for intensities around the caudate region. The classic-BSI intensity normalisation (dividing the intensity by the mean brain intensity) is firstly applied, followed by k-means clustering within the caudate region (dilated by three voxels) to define scan-specific GM-CSF and GM-WM windows. As linear regression is not used to normalise the intensities only a forward CBSI is required.

4.3 Diffusion MR Image Analysis

4.3.1 Pre-Processing

Quality Control (QC)

Diffusion data is split into volumes containing data acquired from different gradient directions, interspersed with B0 images. These B0s are acquisitions with no diffusion sensitivity and therefore act as the structural reference image for the diffusion data. There are different approaches available to QC diffusion data, with some research groups including all data in analyses, some excluding certain corrupted gradient directions only and others discarding whole scans where several directions are affected.

For this thesis all raw diffusion volumes were visually inspected for quality. Missed slices, signal drop-out and coverage issues were recorded. If the anatomy appeared unusual previous time-points were referred to, to ascertain whether this was due to image distortion. As diffusion imaging uses rapid echo planar acquisitions to make acquisition time clinically acceptable, this method is particularly prone to susceptibility artefacts caused by differences in magnetic field susceptibility between tissues, e.g. air in the sinuses, bone and brain-matter. This often causes hyper-intensities around the nasal sinus and geometric distortions in frontal regions (see Figure 4-6). These issues were generally endemic and were only recorded when particularly severe.

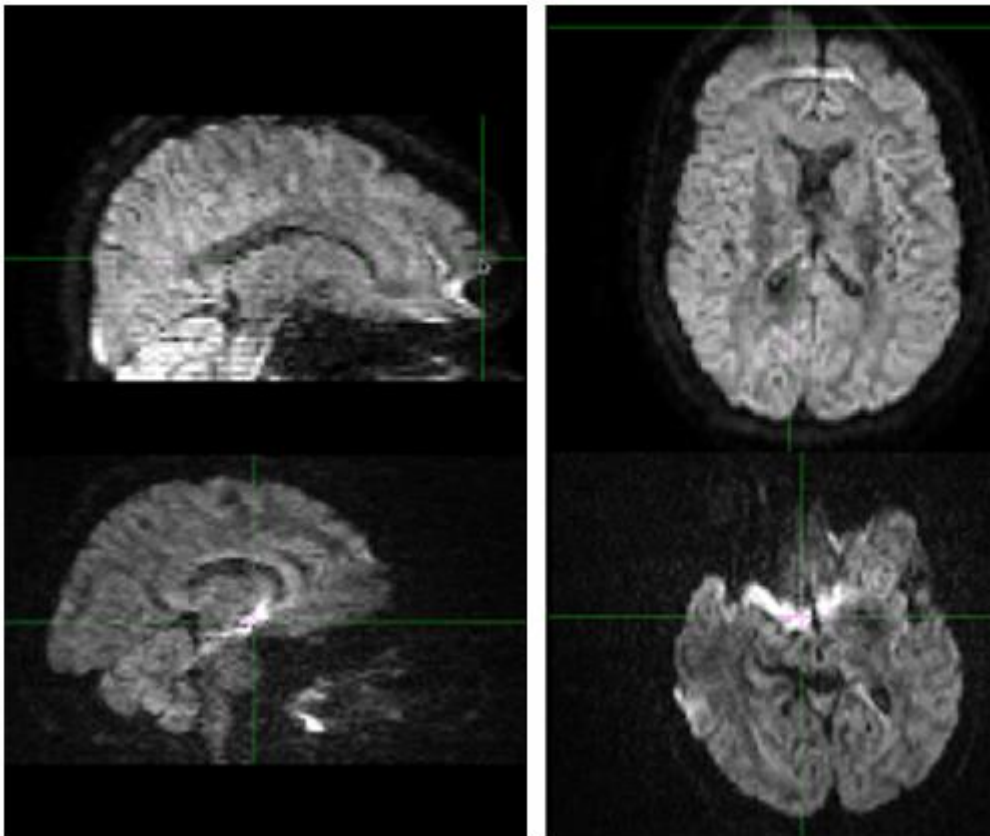


Figure 4-6. Examples of frontal distortions in raw diffusion data (top) and hyper-intensities in regions proximal to the nasal sinuses (below).

In this thesis if scans were deemed to contain significant levels of noise, in multiple gradient directions (enough to corrupt any useful true signal), these were excluded from all analyses. This decision was made, over removal of specific corrupted volumes, in order to avoid any potential disease-related bias introduced by the possibility of removing more data/directions from HD group diffusion scans (due to increased motion artefacts) than the control group.

Motion and Eddy-Current Correction

Eddy currents (localised electrical currents) in the gradient coils induce stretches and shears in the diffusion-weighted images. These distortions are different for different gradient directions. Eddy current correction minimises distortion and head motion artefacts by applying affine or linear registration of all volumes to a reference image. For the PADDINGTON diffusion scans diffusion-weighted images were pre-processed with an initial affine registration to the B0 reference image (or an average of several B0s depending on study site acquisition) to correct for motion and eddy current distortions, and the gradient vectors were updated accordingly. This was done using the Camino software package (www.cs.ucl.ac.uk/research/medic/camino).

Fitting the Tensors

The classic method for fitting tensors to the diffusion data is by means of linear or ordinary least squares regression, such that the sum of squared differences is minimized. This model can fail due to noise or signal

drop resulting in biologically implausible negative eigenvalue estimations (Niethammer et al. 2006). In the PADDINGTON data, using ordinary least squares regression to fit the tensors, 1-2% of voxels showed negative eigenvalues. An estimation method that constrains the eigenvalues to positive numbers is preferable. This was achieved by the use, instead, of non-linear least squares regression.

4.3.2 ROI Analysis

Cross-sectional diffusivity metrics averaged within ROIs are reported in Chapters 13, 14 and 18. Specific methods are reported in these chapters. The diffusion pipeline developed for the large, cross-sectional and longitudinal PADDINGTON study analysis (Chapter 18) is described below.

FA, MD, RD and AD were computed over four ROIs; the WM, CC, caudate and putamen. Firstly, these regions were defined on the T1-weighted images for each individual and saved as binary masks. To reduce partial volume effects (PVEs) in the diffusion metrics, all masks were eroded by one voxel in T1-space. This was preferred to applying a threshold, e.g. an FA cut-off, to avoid circularity between region definition and outcome variable.

The eroded ROIs were then transformed into 'FA-space' using NiftyReg (<http://sourceforge.net/projects/niftyreg>). NiftyReg used a global affine initialisation step (Ourselin et al. 2001) to register the T1-weighted image to the FA image, followed by a non-linear, free-form deformation registration step to improve local alignment (Modat et al. 2010). The transformations generated during these registrations were then applied to the eroded ROIs using tri-linear interpolation. Averaged diffusion metrics were calculated over these regions using the *fsstats* utility within the FSL toolbox (Smith et al. 2004). This process is illustrated in Figure 4-7.

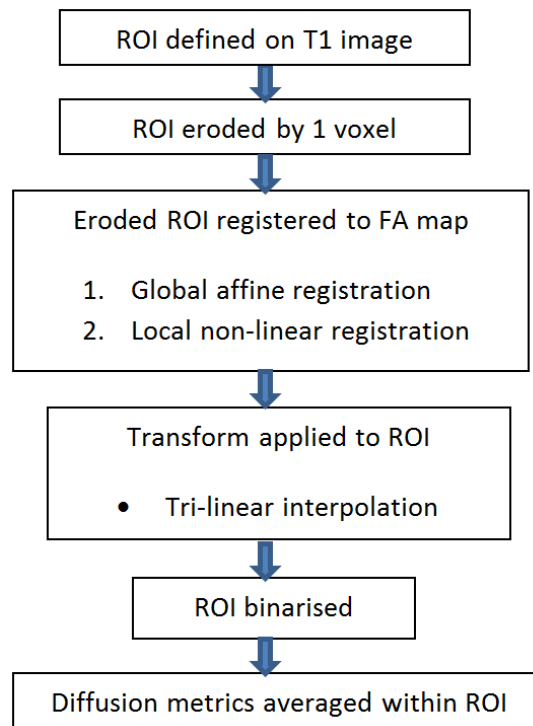


Figure 4-7. An overview of the cross-sectional PADDINGTON diffusion analysis pipeline.

Since the diffusion-weighted sequence did not include full brain coverage (for some participants the cerebellum was only partially covered), the ‘global’ WM region included cut-offs to ensure anatomical consistency in voxels sampled between participants and sites. In brief, the whole-brain regions were thresholded using pre-defined intensity thresholds to exclude GM and CSF voxels. An inferior cut-off excluded all voxels inferior to the orbito-frontal WM. To reduce PVEs, the region was eroded by one voxel in T1-space and further optimised by masking with the caudate and putamen segmentations and an automated GM mask, generated using the expectation-maximisation algorithm LoAd (a locally adaptive cortical segmentation algorithm (Cardoso et al. 2011)) in NiftySeg (<http://niftyseg.sf.net>).

4.3.3 Longitudinal ROI Analysis

For the longitudinal analysis, in order to avoid asymmetrical registration bias a common longitudinal mask was generated for each region by defining a subject-specific ‘half-way-space’. An adapted version of the FSL SIENA application (Smith et al. 2001) was used to affine register the baseline T1 to the follow-up image and vice versa and then calculate the symmetric ‘average’ transformation between the two time-points. Using these two symmetric transformations, a mid-point (i.e. the half-way-space) was defined and then transformation to this space applied to native-space images and their corresponding regional masks.

With the binary regions from baseline and follow-up now in common space the two were multiplied together, in order to ‘zero’ (i.e. remove) any voxels not present at both time-points, only retaining those common to both baseline and follow-up. The same ‘half-way-space’ procedure was then applied to the B0 reference images from the baseline and follow-up, thus defining a common longitudinal diffusion space for

each participant. FA and diffusion maps for each time-point were also transformed into this space. Using NiftyReg, the half-way T1 images were then warped into half-way 'FA-space', using the previously mentioned global affine (Ourselin et al. 2001) and local non-linear registrations (Modat et al. 2010).

Finally, using the inverse of the B0 native-to-half-way-space transformation, the T1 images were moved into native diffusion space for each of the baseline and follow-up. However, to avoid the loss of information due to multiple interpolations, these transformations were composed together using NiftyReg, and applied together in one step to register the T1 half-way-space common mask regions to the native diffusion space, using nearest neighbour interpolation. This process is illustrated in Figure 4-8.

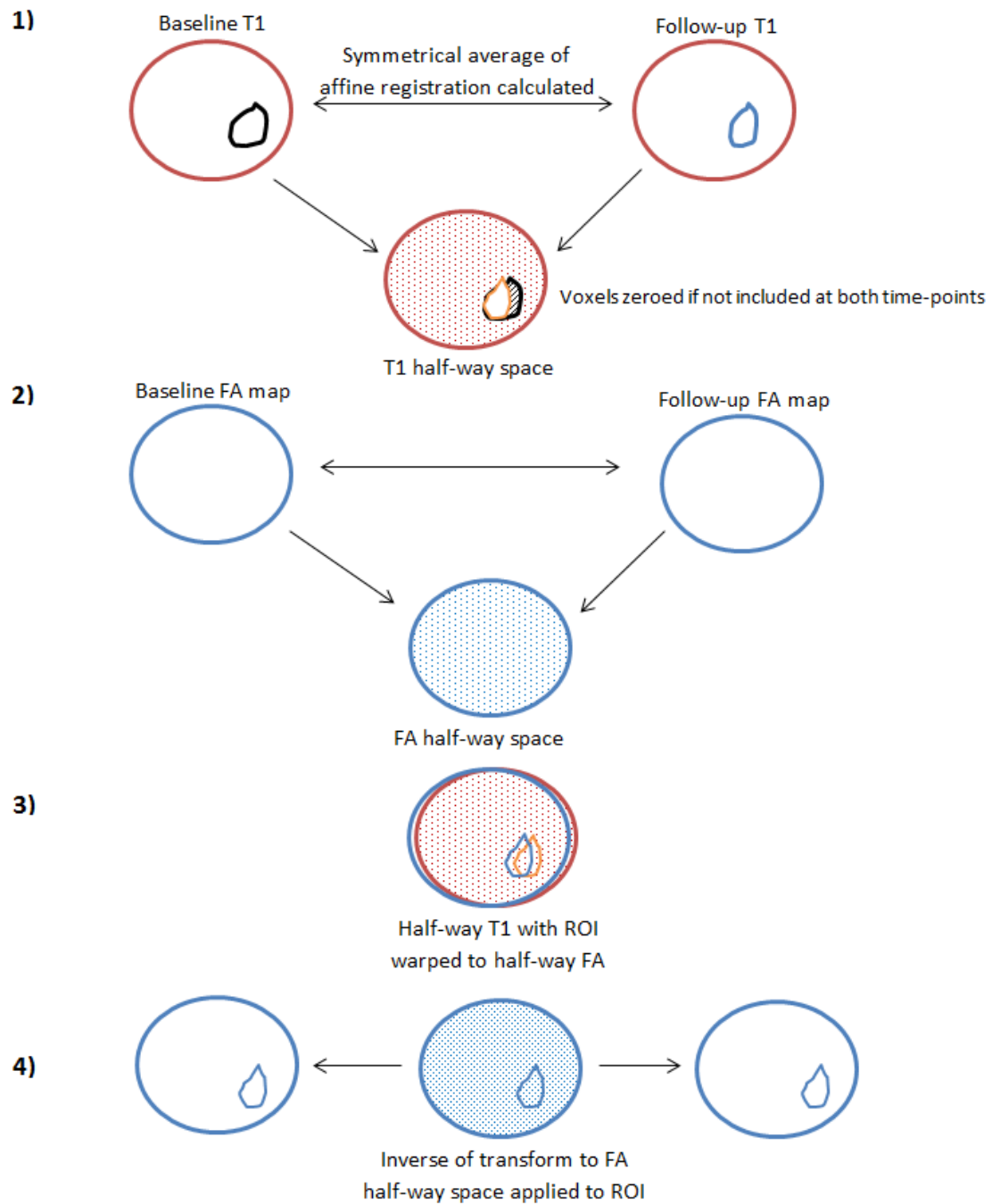


Figure 4-8. An overview of the longitudinal PADDINGTON diffusion analysis. 1) The baseline T1 was affine registered to the follow-up image and vice versa and then a symmetric 'average' transformation was calculated between the two time-points. With the binary regions from baseline and follow-up now in common space the two were multiplied together, in order to 'zero' (i.e. remove) any voxels not present at both time-points, only retaining those common to both baseline and follow-up; 2) the same process was applied to the FA maps; 3) the T1 half-way images and ROI were then warped into FA half-way space; 4) using the inverse of the B0 native-to-half-way-space transformation, the T1 images were moved into native diffusion space for each of the baseline and follow-up and regional means for FA, MD, RD and AD were generated within these regions.

As a further measure to reduce PVEs in the WM region, the original baseline and follow-up T1 images were tissue classified using the LoAd algorithm (Cardoso et al. 2011) to generate GM segmentations. The GM

segmentation was then binarised and transformed to corresponding native diffusion space as above. Any overlapping voxels between the transformed manual WM region and LoAd segmented GM region were excluded.

Thorough visual QC was employed throughout to check for the accuracy of the registrations. Once the manual T1 regions had been transformed to the corresponding diffusion space, regional means for FA, MD, RD and AD were generated using the FSL utility, *fs/stats*.

5 Thesis Methods: Statistical Analyses

All statistical analyses were performed in STATA-IC 11.

5.1 Regression

Generalised least squares regression was used to model between-group (control versus HD) differences for continuous variables e.g. regional brain volume. This model is appropriate as it allows for differing variance between groups and there is often hypothesized to be more variability in certain measures in the HD group than controls and increased variance with advancing stage (Tabrizi et al. 2013). Disease-related associations between imaging metrics and clinical/cognitive performances were computed in the HD group only, by fitting linear regression models.

5.2 Covariates

Covariates are the variables which may confound the statistical comparison under review. All between-group analyses in this thesis were adjusted for age, gender and study site. In longitudinal investigations the scan interval was accounted for. Cross-sectional between-group analyses of regional volumetrics were additionally adjusted for TIV to control for inter-subject variability in head-size, as recommended by Barnes et al. (Barnes et al. 2010).

5.3 Effect and Sample Size Calculations

Effect sizes (ES) are unit-free measures that allow comparisons across methods. They are calculated as the estimated absolute adjusted mean difference of the metric between the HD and control groups, divided by the estimated residual SD of the HD group. These are reported with bias-corrected and accelerated bootstrapped 95% CIs based on 2000 replications (Carpenter and Bithell 2000). An effect size of one implies that the mean change in the HD group is one SD away from that in controls.

Such effect sizes (when squared) are inversely related to sample-size requirements for clinical trials. This association holds under the reasonable assumption that a 100% effective treatment will reduce the mean rate of change in HD cases to that in healthy controls without affecting the variability in these rates. Sample sizes necessary to detect 50% and 20% differences in longitudinal change between groups with 90% power were estimated for this thesis using the standard formula based on two-group Z tests (Julious 2009). The relationship between effect sizes and samples sizes for different assumed therapeutic efficacies, with differing levels of statistical power, are shown in Figure 5-1.

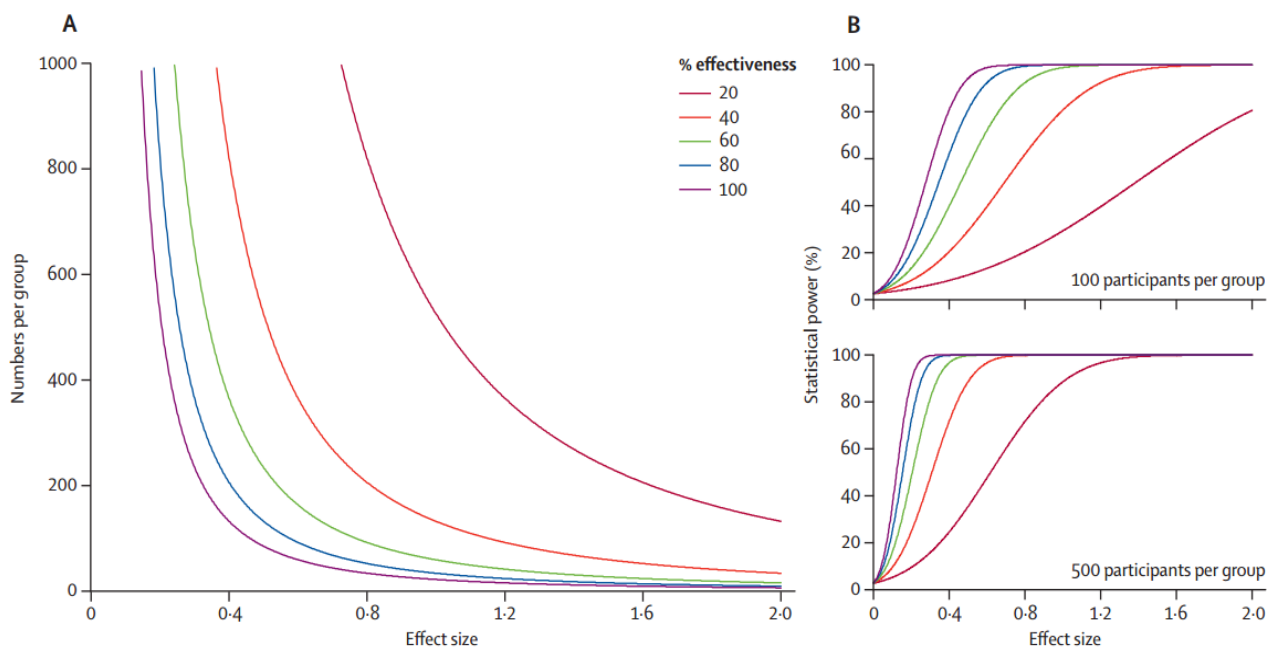


Figure 5-1. Sample size and statistical power calculations for different effect sizes (taken from Tabrizi et al. (Tabrizi et al. 2012)).

- A. Relationships between sample size and effect size for a variable at 90% statistical power for hypothetical treatments ranging from 20% to 100% effectiveness, for a two-year two-group trial.
- B. The relationship between statistical power and effect size for a sample size of 100 participants per group (top) and 500 participants per group (bottom), for a two-year two-group trial.

Both PREDICT-HD and TRACK-HD have provided estimates for sample size requirements in future clinical trials in order to sufficiently power detection of disease-related change using a wide range of biomarkers in different stages of HD (Aylward et al. 2011; Paulsen et al. 2014; Tabrizi et al. 2012). These results and recommendations will be described and discussed in more detail in Part 3 of this thesis.

5.4 Correction for Multiple Comparisons

Statistical comparisons are based on the strength of a finding with, typically, 95% confidence; therefore there is a 5% risk of detecting a false-positive result. Where there are multiple such comparisons undertaken (e.g. mass univariate imaging techniques such as VBM and cortical thickness maps) this risk is greatly increased. Random Field Theory is used in the two most common methods of correcting for multiple comparisons: False Discovery Rate (FDR) and Family-Wise Error (FWE).

Both these methods attempt to assign an adjusted (reduced) p-value to each statistical test to remove false-positive results. FDR correction aims to remove the expected proportion of false discoveries from the significant findings by adjusting the p-value threshold for all tests based on the estimated number of false-positives in the significant results. A FDR correction at 0.05 means that 5% of the 'significant' findings would be false; this is a much smaller quantity than 5% of all tests undertaken. FWE (i.e. Bonferroni inequality) correction is more stringent as it seeks to reduce the probability of even one false discovery by adjusting

the p-value threshold (e.g. 0.05) according to the total number of statistical tests undertaken (not just the significant results). Where possible in this thesis FWE correction is used to ensure as much confidence as possible in statistically significant findings.

6 PART 1. Development, Evaluation and Clinical Application of Tools Sensitive to HD-Related Pathology

At this stage of research development there is a requirement for several further methodological advancements:

- 1 Evaluation of automated alternatives to current manual and semi-automated gold-standard ROI volumetric biomarkers, in order to facilitate large-scale analysis in upcoming clinical trials.
- 2 Continued growth of our knowledge of HD neuropathology by the development of methods for the analysis of under-investigated brain regions in HD; such as the cerebellum.
- 3 Assessment, evaluation and clinical application of an emerging mass-univariate, whole-brain biomarker in HD: FreeSurfer's cortical thickness software.
- 4 Enhancement of current understanding of the HD cognitive phenotype and its underlying neuropathology by development of novel cognitive tests.

Methodological developments, evaluations and clinical applications within these four areas will be conducted in Part 1 of this thesis.

7 Development, Evaluation and Application of Tools Sensitive to Striatal Atrophy in HD

7.1 Background

The striatum, which consists of the caudate and putamen, is a central structure in HD neuropathology (Vonsattel & DiFiglia 1998). As such, volumetry of this region is a strong biomarker candidate for future efficacy trials of potentially disease-altering drugs or therapies, to provide end-points quantifying disease attenuation.

The most sensitive neuroimaging biomarker to emerge from the TRACK-HD study, based on sensitivity to changes early in the disease process and effect sizes, was the CBSI (Tabrizi et al. 2011;Tabrizi et al. 2012;Tabrizi et al. 2013). The CBSI is semi-automated and has previously been validated against a gold-standard manual measure (Hobbs et al. 2009). The pipeline requires segmentation of the baseline caudate; this is performed manually and takes approximately one hour per participant. Large cohorts will be required to sufficiently power clinical trials in HD to detect treatment efficacy. Therefore there is a need for identification of robust fully automated methods, comparable to this semi-automated gold-standard, which would facilitate large-scale volumetric analysis.

There is currently no widely accepted gold-standard method for putamen volumetry. Atrophy within the putamen is thought to be very similar to, if not faster, than caudate atrophy in HD (Aylward et al. 2004;Georgiou-Karistianis et al. 2013b;Halliday et al. 1998) but volumetric methodologies for the putamen are not as sensitive to this change as, for example, the CBSI is for caudate atrophy. This is most likely due to poorer border definition making delineation harder. Manual delineation based on intensity thresholds and careful editing at the voxel-level by eye is widely accepted to be the closest method to a gold-standard available, the assumption being that this method uses all visible structural information in the scan, human knowledge and judgement to delineate the structure. In order to provide a gold-standard against which to evaluate automated methods, a manual delineation protocol was developed in MIDAS software; the process of which is described in Section 7.2.

Method comparisons will facilitate the evaluation and validation of automated techniques to detect longitudinal change in neurodegenerative disease, with the aim of identifying a fully automated alternative, with comparable reliability and sensitivity to the gold-standard measures. The fully automated analyses under evaluation are BRAINS, FIRST, FreeSurfer and STEPS. The results of this method comparison are reported in Section 7.3.

7.2 Development of a Manual Delineation Protocol for the Putamen in HD

7.2.1 Selection of Intensity Thresholds

Initial visual assessment of voxel intensity thresholds, across scanner types and groups, suggested lower and upper threshold limits of 90% and 112% of mean brain intensity respectively for delineation of the putamen. These were appropriate for the majority of scans but excluded some lighter GM voxels and included some darker WM voxels, e.g. Figure 7-1 a) and b) respectively.

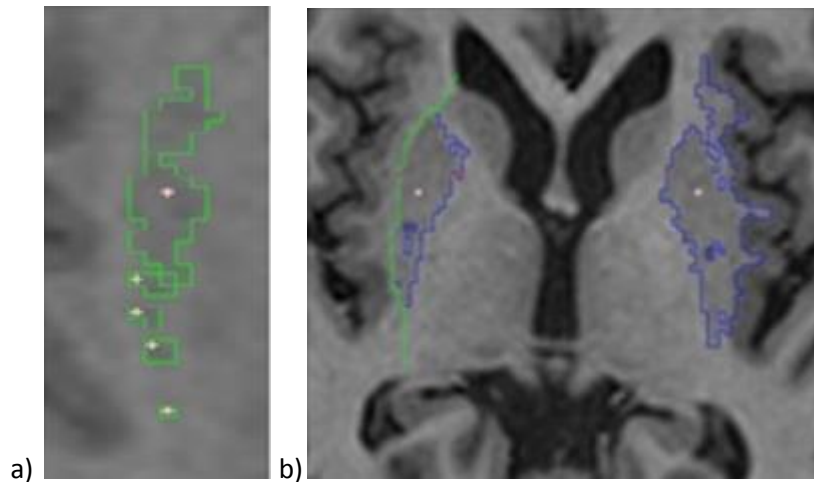


Figure 7-1. Putamen segmentation protocol examples of the initially thresholding stage: a) Some lighter GM voxels may be excluded by the intensity thresholds and therefore several seeds may be needed to roughly fill the structure on each slice; b) Darker WM voxels may also be included and therefore manual editing is often required down the lateral border (example in green), disconnecting the putamen from the WM and claustrum.

7.2.2 Definition of Inferior Cut-off

The inferior putamen is not clearly distinguishable from the bordering nucleus accumbens and inferior WM. It was therefore decided to define a cut-off based on reproducibility rather than anatomical accuracy, i.e. a cut-off based on anatomical landmarks in a higher slice of the putamen than would be anatomically accurate in order to include the whole structure. The anatomical landmark chosen was the WM of the internal capsule which separates the anterior putamen from the caudate (indicated by the red arrow in Figure 7-2). The border of the putamen was judged to be relatively well defined until this point therefore the protocol stipulated that the structure should be delineated until the last slice in which a clear divide between putamen and caudate is seen. Examples of this slice are shown on the left in Figure 7-2. The corresponding subsequent slices below this are shown on the right. This cut-off may be on different slices for the left and right putamen.

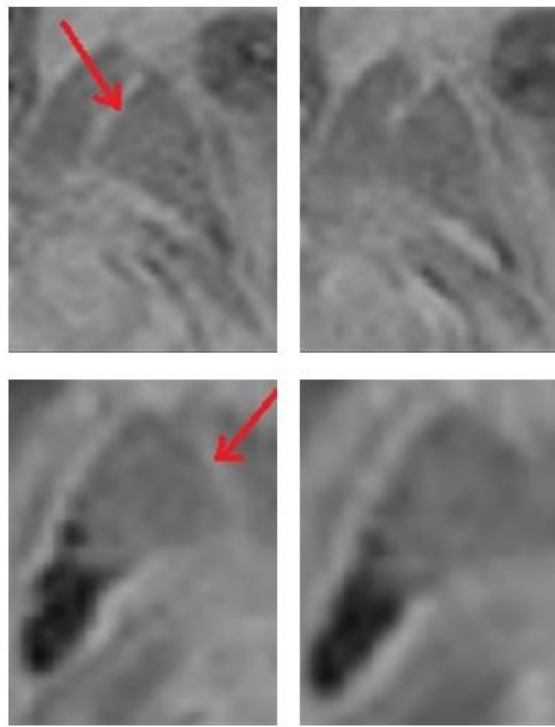


Figure 7-2. Putamen protocol examples: the most inferior slice in which the putamen should be segmented (left) – here the putamen is still clearly separated by WM from the caudate (red arrows). Once this separation is no longer clear (e.g. images on the right) stop the segmentation and do not segment this slice.

Additional rules were required at the lower levels of the putamen: if the tail and head of the putamen are connected (not disconnected by WM or blood vessels) segment the whole structure. If a disconnection is apparent only segment the head (Figure 7-3); dark vessels should be included if they are within the body of the putamen or are continuous with its border but excluded if not. This instruction is for reproducibility reasons and assumes that the amount of hypointensity within the putamen is not clinically meaningful. No investigation was conducted however to assess whether these increase with age, HD or vascular risk factors. If this is the case this protocol may slightly overestimate volume in subjects with vascular or neuropathology.

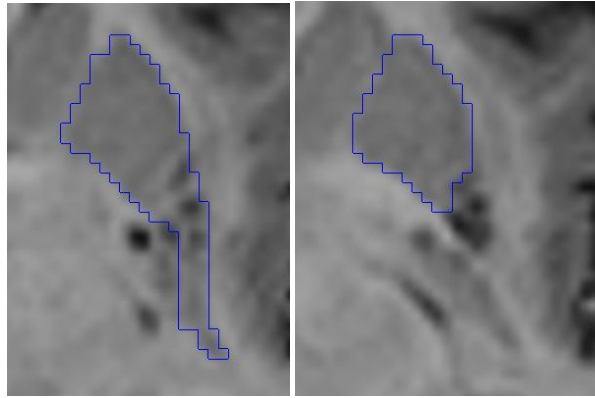


Figure 7-3. Putamen protocol example: if the tail and head of the putamen are connected on that slice segment the whole structure (left). When this becomes split by WM deconvolution the tail (right). Dark spots within the structure (e.g. in left-hand image) should be included.

It was also defined that the putamen segmentation in both the axial and sagittal views must have smooth biologically plausible outlines (e.g. Figure 7-4).

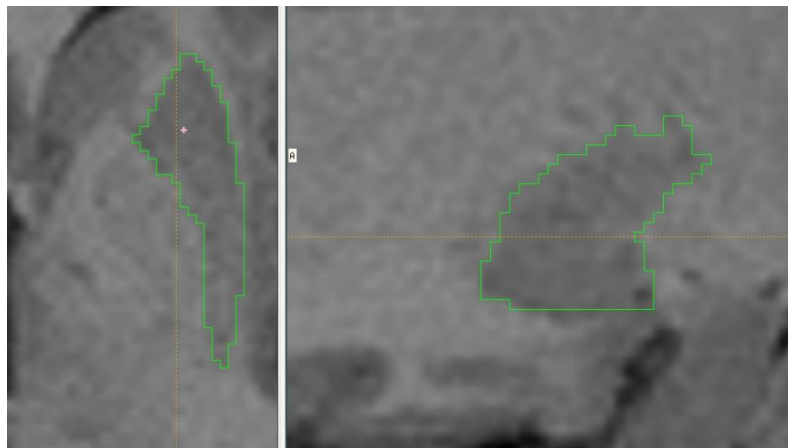


Figure 7-4. Putamen protocol example: ensure that edges are smooth and biologically plausible in both views.

7.2.3 Reproducibility

For a structure such as the putamen the reproducibility of volume estimates would ideally be >99% (i.e. less than 1% difference in volume estimates calculated as $(A-B)/\text{mean} \times 100$), on the basis that changes over time are small and subtle, and therefore in small regions such as the putamen more than 1% measurement error may mask these changes. Scans from six participants from the TRACK-HD study (two controls, two preHD and two HD participants; not included in the protocol development process) were analysed by two trained analysts and repeated after a week interval to assess inter- and intra-analyst protocol reproducibility. The inter-analyst variability in volume estimates was on average 3.6% (0.7% to 5.2%). The intra-analyst repeat variability in volume estimates was on average 3.5% (0.2% to 6.8%). The two scans with the highest percentage differences in putamen volume estimates contained notably smaller putamens than the other scans, thus increasing the percentage change for similar differences in ml. This bias towards greater proportional inaccuracy in atrophied brains is likely to introduce a systematic difference between controls and HD subjects. This level of reproducibility is not to the desired level of accuracy. The cause of

this is most likely the poor border definition around the putamen and reflects the increased difficulty of segmenting this region compared with other regions such as the caudate and whole-brain. It should also be noted that this reproducibility test was limited by sample size.

7.2.4 Inter- and Intra-Scanner Type Tests

Background

Subtle differences in scan acquisition, such as varying tissue intensities and tissue contrast, may affect the reliability and consistency of segmentation protocols. It was therefore necessary to test whether this protocol was consistent across scanner types (Siemens vs Philips) in both control and HD groups and also between Siemens scanner subtypes: Tim Trio vs Verio vs Allegra. Groups were analysed separately in order to detect any potential disease bias in the measurements.

Cohort

Baseline scans from a sample of 44 PADDINGTON study participants (22 controls and 22 HD individuals; not included in the protocol development process) were selected for analysis in this investigation of protocol performance across scanner types and subtypes. By design, these participants were split between study sites and scanner types, and balanced as far as possible in terms of age, gender and disease burden. Sample demographics are shown in Table 7-1.

Image Acquisition

Scans were acquired at four study sites. One of these sites (Leiden) used a Philips (Achieva) 3T scanner. The other three sites (London, Paris and Ulm) used Siemens 3T scanners: Tim Trio, Verio and Allegra respectively.

Image Analysis

All scans were analysed using the newly developed manual delineation protocol for the putamen; full protocol in Appendix Section 23.4. Each delineation required approximately one hour. All segmentations were visually assessed for quality.

Statistical Analysis

Regression models were fitted to: a) detect significant between-site group differences in demographics; b) assess differences in putamen volume estimates between groups and scanner types, adjusting for age, gender and TIV. Variance-comparison tests were used to quantify between-method differences in estimate variability.

Results

Table 7-1. Between-scanner test of the putamen segmentation protocol - sample demographics.										
	Controls					Early HD				
Scanner type	Philips (Achieva)	All Siemens	Siemens Tim Trio	Siemens Verio	Siemens Allegra	Philips (Achieva)	All Siemens	Siemens Tim Trio	Siemens Verio	Siemens Allegra
N	10	12	4	4	4	10	12	4	4	4
Gender: M/F	6/4	6/6	2/2	2/2	2/2	2/8	6/6	2/2	2/2	2/2
Age (years):	49.4	53.4	51.7	57.8	50.6	52.6	55.2	49.5	57.8	58.3
Mean (SD)	(5.8)	(5.9)	(6.0)	(5.1)	(4.9)	(8.1)	(7.3)	(4.2)	(9.1)	(5.4)
Disease Burden*	NA	NA	NA	NA	NA	386.8 (75.3)	390.0 (79.2)	444 (59.7)	324.1 (56.0)	401.8 (80.4)
*Disease Burden derived from the formula by Penney et al. (Penney et al. 1997). NA = not applicable.										

The sample was similar between groups (HD and controls) and scanner types and subtypes (Table 7-1).

There was however a significant difference in disease burden in the HD groups scanned on the Siemens Tim Trio and Siemens Verio scanners ($p=0.026$) and significant differences in age between the Philips and Verio control groups ($p=0.028$) and Tim Trio and Verio HD groups ($p=0.042$).

Visual analysis of the scans did not highlight any problems with protocol performance across scanner types; the thresholds appeared to respond similarly on all image acquisitions. There was however a suggestion, from the experience of analysing the scans, that more atrophied regions were more difficult to delineate than healthy-sized putamen. This was thought to be due to a degrading of the border contrast and a lightening of the putaminal GM.

There was a significant difference in volumes (across all scans) between groups (HD - controls; -2.79ml (95% CI $-3.55, -2.03$) $p<0.001$). The raw putamen volumes, split by group and scanner type, are shown in Figure 7-5A. After adjustment for between-site differences in age, gender and TIV, putamen volumes were found to differ significantly between scanner types in the control group ($p=0.044$); with Philips scans outputting significantly smaller volume estimates ($-.96\text{ml}$ (95% CI $-1.90, -.03$)). There was no significant between-scanner difference in the HD group ($-.09\text{ml}$ ($-1.39, 1.21$) $p=0.887$), although the scanner means showed a similar trend for smaller volumes generated from the Philips scans.

With both scanner types the variance in the HD group was significantly greater than that in the control group ($p<0.0001$). Volume estimates in both groups showed significantly larger variance from the Siemens scans compared with the Philips scans ($p<0.001$).

The raw putamen volumes, split by group and scanner subtype, are shown in Figure 7-5B. This plot highlights larger volume estimates in the HD group scanned by the Siemens Verio scanner compared with all other scanners. This (unadjusted) difference was significant compared with Tim Trio ($p=0.025$) and

Allegra ($p=0.019$) scan volumes. When adjusted for age, gender, TIV and disease burden however these differences were no longer significant ($p=0.299$ and $p=0.406$).

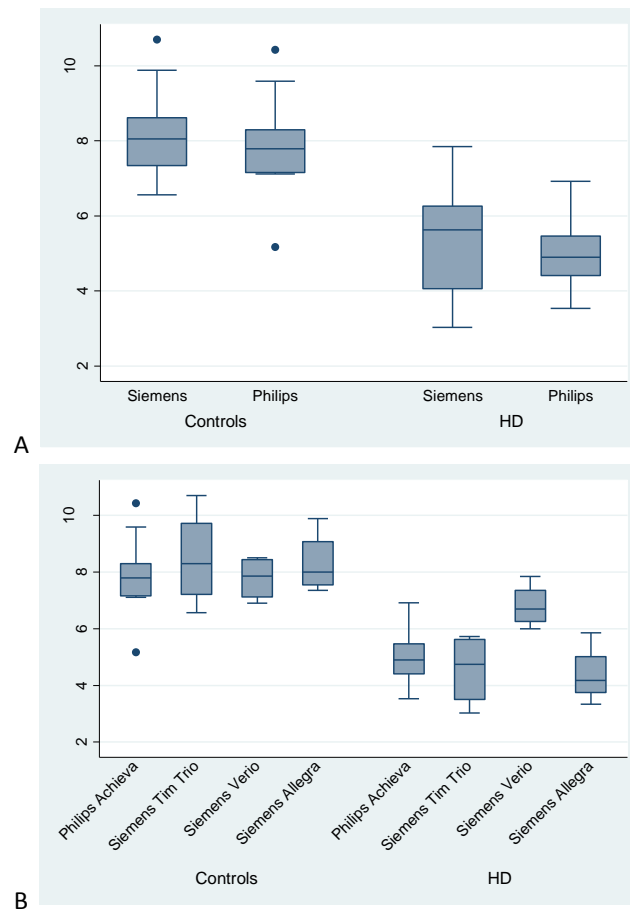


Figure 7-5. A) Raw putamen volume estimates in each group separated by scanner types; B) Raw putamen volume estimates in each group separated by scanner types and subtypes.

Discussion

This was an investigation of inter- and intra-scanner type variability in volume estimates outputted from the newly developed protocol described in this chapter. Strong between-group differences were detected. Between-scanner differences were also present but to a much smaller degree. Variability in volume estimates between Siemens scanners appeared to be as variable as when compared with Philips scans. These differences do not suggest the presence of any large systematic bias between scanner types or subtypes, and are thought to be typical of multi-site data.

There was a suggestion, from experience and visual inspection, that measurement errors were higher in the HD group compared with the control group due to degradation of the boundary contrast on the T1-weighted image of scans showing more severe atrophy; and consequently increasing the difficulty in delineating these atrophied structures. This may be reflected by the significantly larger volume estimate variance in the HD group compared with the control data. Alternatively this may be a biologically accurate

finding as there is often hypothesized to be more variability in certain measures in the HD group compared with controls and increased variance with advancing stage (Tabrizi et al. 2013).

In the control group the putamen volume estimates from the Siemens scans were significantly larger than those from the Philips scans; a trend was also seen in the HD data, although reduced. This result in the control group was made significant by an outlier in the Philips data with a very small volume estimate (a 56 year-old female with a small TIV (the 5th smallest out of 44)). When this outlier was removed this difference was no longer significant (-.76ml (95% CI -1.77, .25) p=0.132).

Estimate variance was significantly larger from the Siemens scans, compared with the Philips scans. This was most likely due to the fact that the Siemens scan volumes were combined across three different Siemens scanner subtypes, whilst the Philips scan volumes were all from the same scanner. The data indicates that the Siemens HD scan volume estimates from the Verio scanner play a large role in this increased variance. The larger volume estimates in the HD group scanned on the Siemens Verio in comparison to the estimates from the other scanners seem however to be due to differences in group characteristics; between-scanner subtype differences were no longer significant after adjustment for age, gender, TIV and disease burden.

This investigation was limited by the impracticality of scanning the same test groups on each scanner; due to the large amount of resources this would require and the unfeasibility of this amount of international travel in participants with HD. The imbalance between scanner types was also a limitation of the data available.

Overall the results from this investigation do not give any strong contraindication for use of this protocol for putamen volumetry across scanner types. Despite a suggestion of minor variability in volume estimates and estimate variance between scanner types and between groups, all these findings can be partially or fully explained by other factors.

7.2.5 Protocol Development: A Summary

As expected, the putamen was found to be a difficult structure to delineate due to the poor border contrast in some areas. This was particularly the case along the inferior border, resulting in the need for a clearly defined anatomical cut-off. More atrophied structures were deemed to be more difficult to delineate than healthy-sized structures owing to degradation in the putaminal GM and consequently the border definition. This potential disease-bias was reflected in the poorer reproducibility in the delineations of smaller structures. Slight between-scanner differences were seen. These effects however were not large enough to compromise the data, especially if groups are well matched between study sites. Despite the suboptimal performance of this protocol in its consistency and reproducibility, manual delineation is still deemed to be the gold-standard methodology. This is based on the assumption that the use of all visible structural

information in the scan, human knowledge and judgement to delineate the structure is more accurate than any other method. Accordingly, this manual delineation protocol (the final version of which is included in Appendix Section 23.4) will be used in the following method comparison, with the caveats regarding measurement error and potential systematic bias taken into consideration.

7.3 A Method Comparison

7.3.1 Aim

With upcoming large clinical trials on the horizon there is a need for the identification of robust fully-automated methods for striatal volumetry with comparable reliability and sensitivity to the manual (putamen) and semi-automated (CBSI) gold-standards. This method comparison aimed to identify the best automated alternatives by evaluating striatal volumetrics outputted from four widely used software packages: BRAINS3, FIRST, FreeSurfer and STEPS.

7.3.2 Methods

Cohorts

The PADDINGTON cohort (n=101) was used for this study; 37/40 controls and 53/61 HD participants were scanned at baseline, 6- and 15-month time-points. For the putamen method comparison a subset of 20 controls and 28 HD participants were selected and analysed. For the caudate method comparison the whole cohort was analysed with the caudate gold-standard methodology (the CBSI). All outputs were visually assessed for quality. Those with 6- and/or 15-month CBSI estimates failing QC were excluded in order to establish a gold-standard which was as close as possible to the reality of the caudate change. Consequently, all data from five controls and five HD participants were removed, resulting in a final sample size of 32 controls and 48 HD participants. Demographics of these samples are show in Table 7-2.

Table 7-2. Demographics of putamen and caudate method comparison samples.				
	Putamen Method Comparison (n=48)		Caudate Method Comparison (n=80)	
	Controls (n=20)	HD Stage I (n=28)	Controls (n=32)	HD Stage I (n=48)
Age: mean (SD), range	51.0 (6.8), 38.3 - 64.2	47.1 (8.9), 26.8 - 63.3	52.1 (8.8), 29.0 - 66.6	49.0 (9.8), 26.8 – 67.3
Gender: F/M	10/10	12/16	19/13	30/18
Site: Leiden/London/Paris/Ulm	5/5/5/5	7/7/7/7	5/10/8/9	12/13/9/14

Image Acquisition

3T T1-weighted MRI data were acquired - full acquisition parameters are detailed in Section 3.4. Data were pseudoanonymised and archived on a secure web-portal. QC was performed on all scans checking for artefacts such as movement and intensity inhomogeneity, and sufficient tissue contrast for analysis. All scans included in this method comparison passed QC.

Image Analysis

N3 bias-corrected (Sled et al. 1998) T1-weighted scans were analysed to output putamen and caudate segmentations using the following methods – more details of automated methods can be found in Sections 4.2.3 to 4.2.7:

1. **Manual Putamen Delineation:** A novel protocol (described above in Section 7.2) was developed using in-house MIDAS software (Freeborough et al. 1997). Analysis was conducted blinded to group and with all three participants' scans analysed in parallel. The protocol used is included in the Appendix Section 23.4. Briefly this involved:

Rigid registration of T1-weighted scans to MNI305 atlas space to facilitate consistent application of landmark-derived cut-offs to be included in the segmentation protocol. Segmentation was initialized using pre-defined intensity constraints set at 90% and 112% of the mean brain intensity of each scan; this created an initial outline of the structure. The anterior and lateral borders of the putamen were defined by the WM of the internal and external capsules respectively, whilst voxel intensity differentiated the putamen from the globus pallidus along the medial border. This outline was refined following these anatomical specifications. Segmentations required approximately one hour.

2. **CBSI (Hobbs et al. 2009):** The caudate was manually delineated on the baseline scan and used to compute the BSI – details in Section 4.2.11. Each baseline segmentation required approximately one hour.
3. **BRAINS3 (Magnotta et al. 2002):** BRAINS software (Magnotta et al. 2002) utilized data from both T1- and T2-weighted scans to segment the putamen and caudate from surrounding tissue. A neural net, trained based on human rater definition of these structures and atlas-based structure identification, was applied followed by an additional boundary correction to ensure no structural overlap.
4. **FIRST (Patenaude et al. 2011):** FIRST uses learned shape models trained with 336 T1-weighted scans manually labelled with 15 subcortical structures, including the putamen and caudate. The models of these structures were fitted to the observed intensities in the T1-weighted images to produce the most probable shape instances given the data. As is default, boundary correction was applied to the caudate but not the putamen segmentations.
5. **FreeSurfer v5.3.0 (Fischl et al. 2002):** FreeSurfer performed segmentations based on probabilistic information automatically estimated from a manually-labelled training set.
6. **STEPS (Cardoso et al. 2013):** STEPS used a library of manual delineations which was referenced to identify the most closely matched putamen and caudate segmentations to the voxel intensities in the T1-weighted image. These were then fused to form an optimal structural outline.

Baseline TIV estimations were also outputted using SPM8's New Segment toolbox. These were used to adjust for inter-subject variability in head-size.

Quality Control (QC)

All segmentations were visually inspected. Those with missing volumes or extreme errors (e.g. incorrect positioning on the scan) were excluded from analyses. Due to the relative unknown of the quality of putamen segmentations from different methods, an additional systematic visual analysis was conducted. Descriptions and rough percentages of each error were calculated and tabulated.

Statistical Analysis

Volumes were extracted from baseline, 6- and 15-month segmentations in ml. The change in volume from baseline over the two time intervals (six and 15 months) was then calculated within each participant as the raw (ml) and, for the between-group comparisons, percentage change $((V2-V1)/V1*100)$.

Raw change estimates from the automated methods were compared, within-group, with the gold-standard measure. The null hypothesis was that both methods being tested would give the same value. Paired t-tests assessed the significance of between-method differences. T-tests were run to assess the significance of volume change from zero. Paired variance-comparison tests report between-method differences in estimate variability and pairwise Pearson's correlation coefficients were used to quantify agreement between methods.

Bland Altman and scatter plots illustrate raw change estimate agreement between automated methods compared with the gold-standard measure. Controls and HD patients are plotted separately to identify any potential disease-related bias.

Between-group differences in percentage volume change, adjusted for scan interval, were tested using generalised least squares regression models, adjusting for age, gender and study site. Effect sizes were calculated as the estimated absolute adjusted mean difference of the metric between the HD and control groups, divided by the estimated residual SD of the HD group. These are reported with bias-corrected and accelerated bootstrapped 95% CIs based on 2000 replications (Carpenter & Bithell 2000).

7.3.3 Results

Putamen Volumetry Method Comparison

Of the putamen segmentations the following were missing:

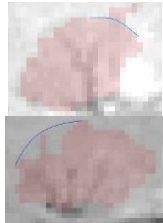
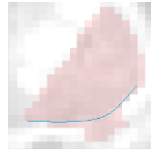
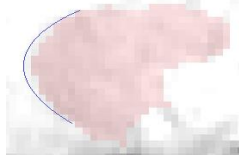
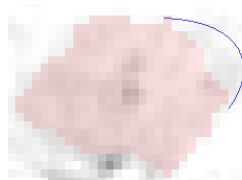
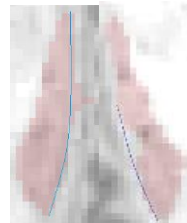


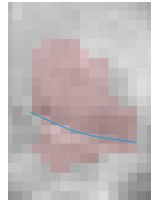
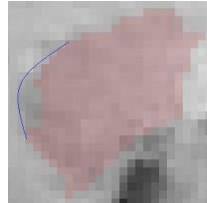
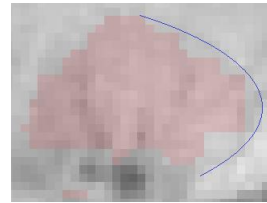

- FIRST: Two segmentations from HD participants and one from a control participant failed QC at baseline. One of these HD participants' segmentations also failed 6-month QC and a separate HD participant's 6-month scan failed to complete the pipeline. One control and one HD participant's segmentations failed 15-month QC. This resulted in one control and three HD participants with

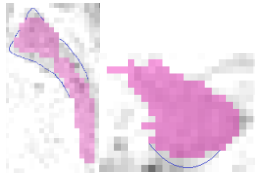
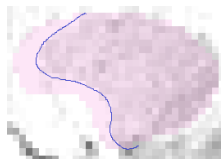
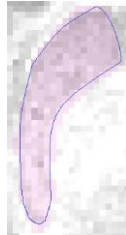
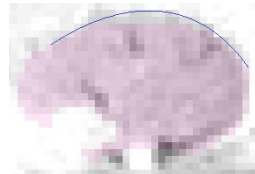



missing 6-month change estimates; two controls and three HD participants with missing 15-month change estimates.

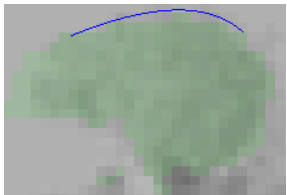
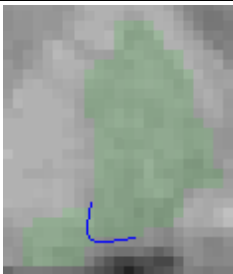
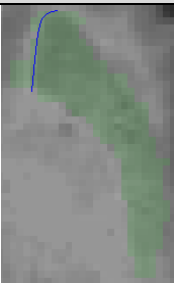
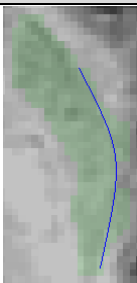
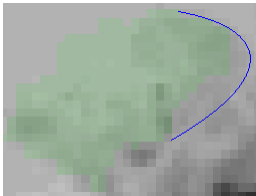
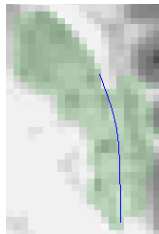

- FreeSurfer: one control participant and one HD participant were missing 6-month and 15-month change estimates due to QC failure on all three serial scans.
- STEPS: one control was missing a 15-month visit volume estimate.

A systematic visual analysis of the automated segmentations highlighted several common inaccuracies when compared against the visible border of the putamen - details of the visual QC results are reported with examples in Table 7-3. Approximately 97% of BRAINS3 segmentations were judged to contain moderate errors, ~66% of those scans (~21% of which were control scans, ~79% of which were HD scans) were deemed to contain severe anatomical errors - most notably, ~38% showed substantial inclusion of inferior WM; FIRST segmentations were typically one voxel too loose along the lateral, medial and posterior borders but tighter to anterior and superior edges, with several common regions of moderate error effecting ~32% of segmentations (~44% of which were control scans and ~56% of which were HD scans) and ~3% of segmentations failing due to poor mask positioning (~44% of which were control scans, ~56% of which were HD scans); FreeSurfer commonly included extensive inferior WM and ~50% of all segmentations exhibited severe lateral leakage (~45% of which were control scans, ~55% of which were HD scans); STEPS delineations were the most anatomically accurate, with minor errors in just ~20% of segmentations (~72% of which were control scans, ~28% of which were HD scans) – the most notable error being exclusion of lateral regions.

Table 7-3. A systematic visual assessment and summary of putamen segmentation errors outputted from different automated methods.

Method	Superior	Inferior	Anterior	Posterior	Lateral	Medial	Other
BRAINS3	Moderate Errors						
		 1-2 slices affected				 1-2 slices affected	No other moderate errors detected
	~20%: moderate leakage/exclusion of superior GM	~27%: moderate inferior leakage	~13%: anterior GM excluded	~50%: posterior GM excluded	~33%: lateral leakage/exclusion	~27%: moderate medial leakage	
	Severe Errors						
		 3+ slices affected			No severe lateral errors detected	 3+ slices affected	No other severe errors detected
	~30%: severe leakage/exclusion of superior GM	~38%: extensive inferior leakage	~17%: larger regions of anterior GM excluded	~27%: significant regions of posterior GM missing		~20%: severe medial leakage	

Method	Superior	Inferior	Anterior	Posterior	Lateral	Medial	Other
FIRST	Common Inaccuracies						
	No common superior inaccuracies		No common anterior inaccuracies				No other common inaccuracies detected
		Inaccurate inferior anatomy for all		Typically loose with extensive WM included	~1 voxel too large along lateral and medial edges in almost all segmentations		
	Notable Errors						
		No notable inferior errors detected			No notable lateral or medial errors detected		
	~15%: superior section missing		~15%: region of anterior section missing	~12%: posterior section missing			~3%: fails due to poor mask positioning

Method	Superior	Inferior	Anterior	Posterior	Lateral	Medial	Other
FreeSurfer (version 5.1.0)	Common Inaccuracies						
				No common posterior errors detected		No common medial errors detected	No other common errors detected
	Commonly tight along superior border	Almost all segmentations show severe inferior leakage	Minor anterior border leakage		Lateral border leakage		
	Notable Errors						
	No notable superior errors detected	No notable inferior errors detected		No notable posterior errors detected			No other notable errors detected
			~5%: regions of anterior GM excluded		~50%: severe lateral border leakage	~9%: sections of medial putamen missing	

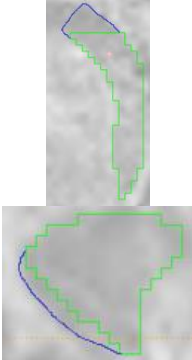
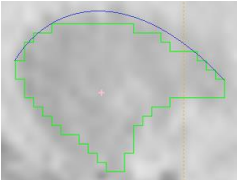
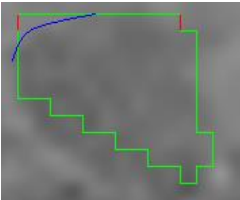
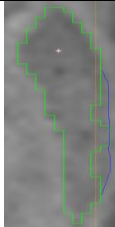
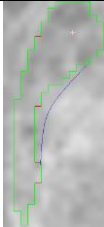
Method	Superior	Inferior	Anterior	Posterior	Lateral	Medial	Other
STEPS	Common Inaccuracies						
	No common superior errors detected		No common anterior errors detected	No common posterior errors detected	No common lateral errors detected	No common medial errors detected	No other common errors detected
		Anatomically inaccurate inferior cut-off but relatively consistent					
	Notable Errors						
		No notable inferior errors detected		No notable posterior errors detected			No other notable errors detected
	~10%: small superior sections missing		~7%: minor inclusion of anterior WM		~20%: small to large sections of lateral GM excluded	~2%: small medial sections missing	
The blue outline highlights the putamen border where segmentations are inaccurate.							

Table 7-4. Summary statistics of raw putamen 6- and 15-month volume change data.

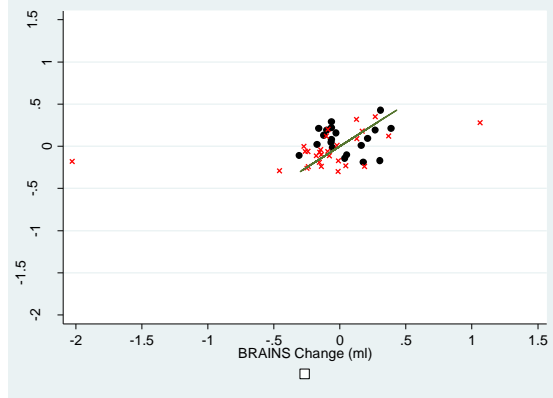
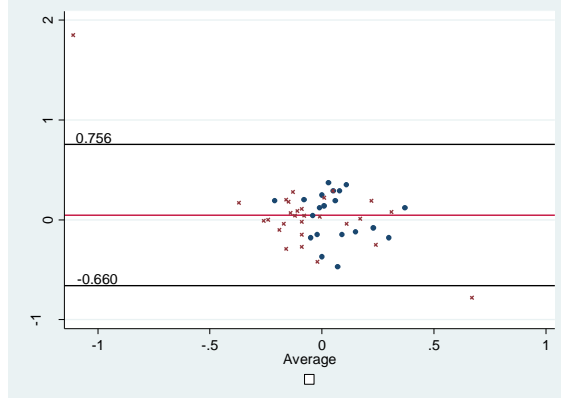
Method	Group	n	Mean (SD), ml	Mean Diff. (95% CI), ml (Manual – Automated)	Ratio of Variance (95% CI) (Manual/Automated)	Pearson's Rho (Manual vs Automated)
6-Month Change						
Manual	Controls	20	.078 (.17)			
	HD	28	-.046 (.19)			
BRAINS3	Controls	20	.036 (.190)	.042 (-.071 , .155) p=0.4465	.874 (.543 , 1.406) p=0.573	0.0853
	HD	28	-.098 (.471)	.052 (-.111 , .215) p=0.5185	.403 (.283 , .573) p<0.001	0.4524
FIRST	Controls	19	.105 (.326)	-.016 (-.145 , .112) p=0.7934	.498 (.332 , .748) p=0.001	0.5793
	HD	25	-.063 (.270)	.018 (-.117 , .153) p=0.7886	.727 (.478 , 1.105) p=0.133	0.0400
FreeSurfer	Controls	19	.056 (.209)	.026 (-.051 , .102) p=0.4948	.814 (.560 , 1.184) p=0.270	0.6625
	HD	27	-.042 (.139)	-.017 (-.082 , .048) p=0.5943	1.289 (.908 , 1.832) p=0.152	0.4928
STEPS	Controls	20	.040 (.197)	.038 (-.033 , .108) p=0.2822	.842 (.586 , 1.210) p=0.341	0.6641
	HD	28	-.033 (.102)	-.012 (-.071 , .046) p=0.6700	1.858 (1.357 , 2.543) p<0.001	0.6110
15-Month Change						
Manual	Controls	20	.018 (.26)			
	HD	28	-.100 (.23)			
BRAINS3	Controls	20	.007 (.203)	.010 (-.140 , .161) p=0.8867	1.300 (.808 , 2.092) p=0.274	0.0673
	HD	28	-.217 (.389)	.118 (-.028 , .264) p=0.1093	.597 (.413 , .864) p=0.007	0.3537
FIRST	Controls	18	.005 (.371)	.040 (-.154 , .234) p=0.6699	.701 (.429 , 1.143) p=0.150	0.2767
	HD	25	-.183 (.310)	.095 (-.028 , .218) p=0.1252	.774 (.530 , 1.131) p=0.181	0.4344
FreeSurfer	Controls	19	.065 (.258)	-.052 (-.187 , .082) p=0.4264	1.047 (.672 , 1.633) p=0.834	0.4424
	HD	27	-.085 (.230)	-.026 (-.104 , .053) p=0.5064	.995 (.726 , 1.364) p=0.975	0.6273
STEPS	Controls	19	.026 (.201)	.012 (-.109 , .134) p=0.8361	1.267 (.806 , 1.991) p=0.296	0.4054
	HD	28	-.096 (.182)	-.004 (-.085 , .078) p=0.9253	1.279 (.910 , 1.797) p=0.153	0.5100
Mean 6- and 15-month change and the SD of these estimates, outputted by each method. Comparisons with the manual method included: paired t-tests to assess the significance of the between-method difference; two-sample variance-comparison tests; and pairwise correlations.						

No control group change estimates, over either interval, were significantly different from zero. In the HD group there were no significant change estimates (different from zero) over six months but over 15 months all methods detected significant atrophy ($p < 0.05$); except the FreeSurfer estimates which were only of borderline significance (-0.085ml (SD 0.230) $p = 0.066$). No method reported significantly different 6- or 15-month change estimates compared to the manual measure although no method correlated with this gold-standard at $\text{Rho} > 0.656$ (FreeSurfer and STEPS showed the highest correlations with this gold-standard). The variance within BRAINS3 change estimates in the HD group was significantly higher than that of the manual measure over both intervals ($p < 0.001$ and $p = 0.007$ respectively). The FIRST 6-month change estimates in the control group also showed significantly larger variance than the manual measure ($p = 0.001$). Conversely, the STEPS 6-month change estimate in the HD group showed significantly lower variance than the manual ($p < 0.001$).

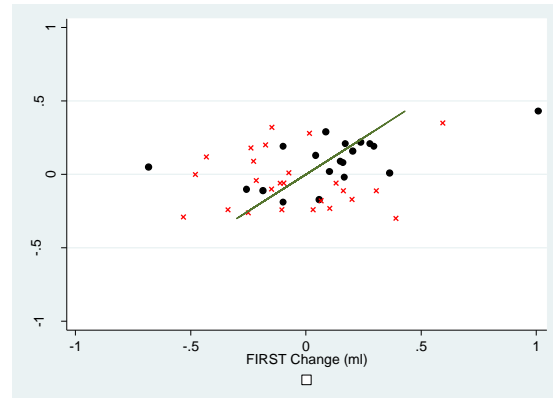
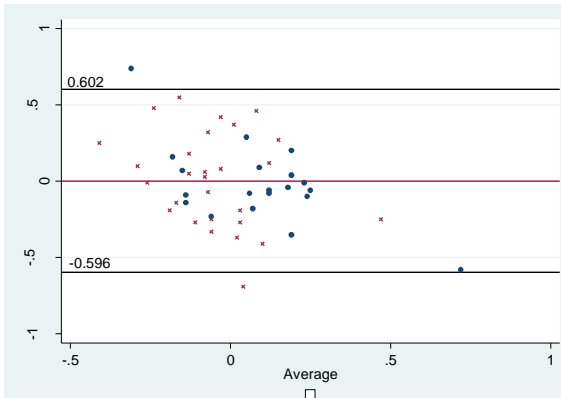
Bland Altman and scatter plots between 6- and 15-month change estimates from the four automated methods and manual delineation are shown in Figure 7-6. Large outliers within the HD group are apparent from the BRAINS3 plots. Whilst the correlations with the manual gold-standard were not tight, there were no clear biases in the measures either across atrophy rates or between groups.

6-Month Putamen Volume Change Estimates versus Manual Gold-Standard

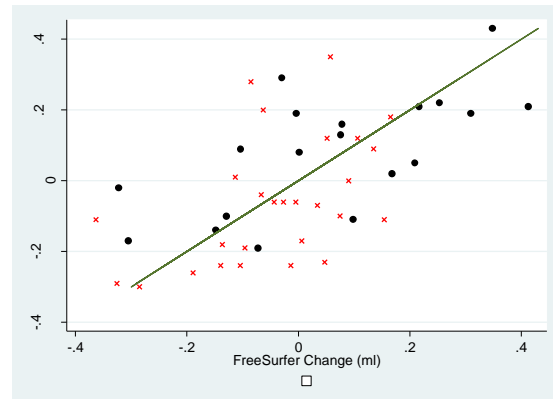
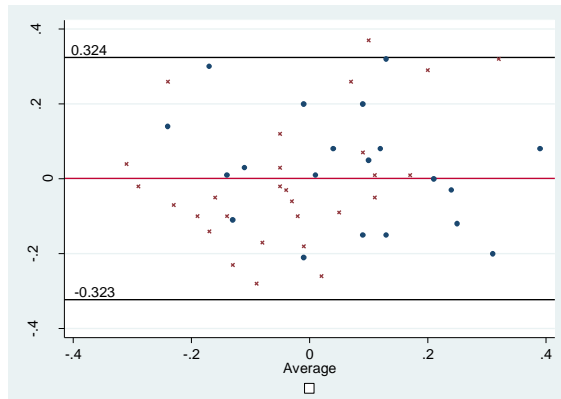
BRAINS3



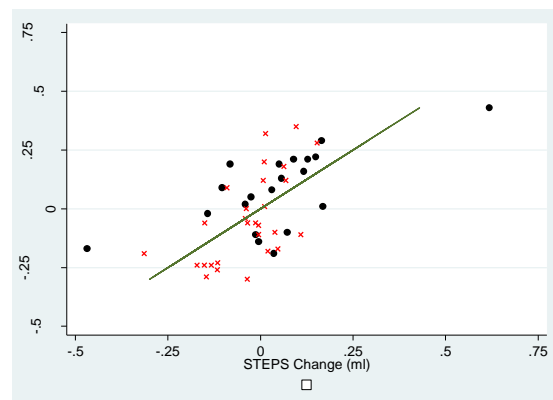
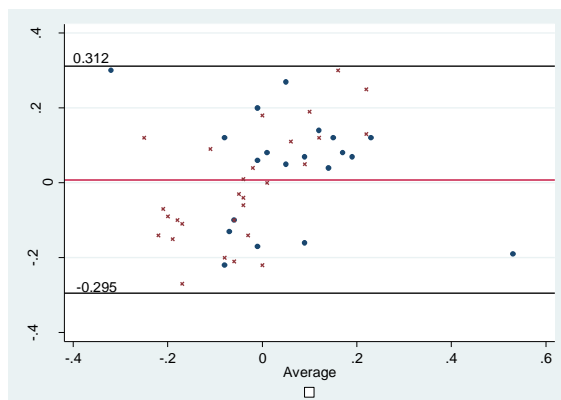
FIRST



FreeSurfer



STEPS



15-Month Putamen Volume Change Estimates versus Manual Gold-Standard

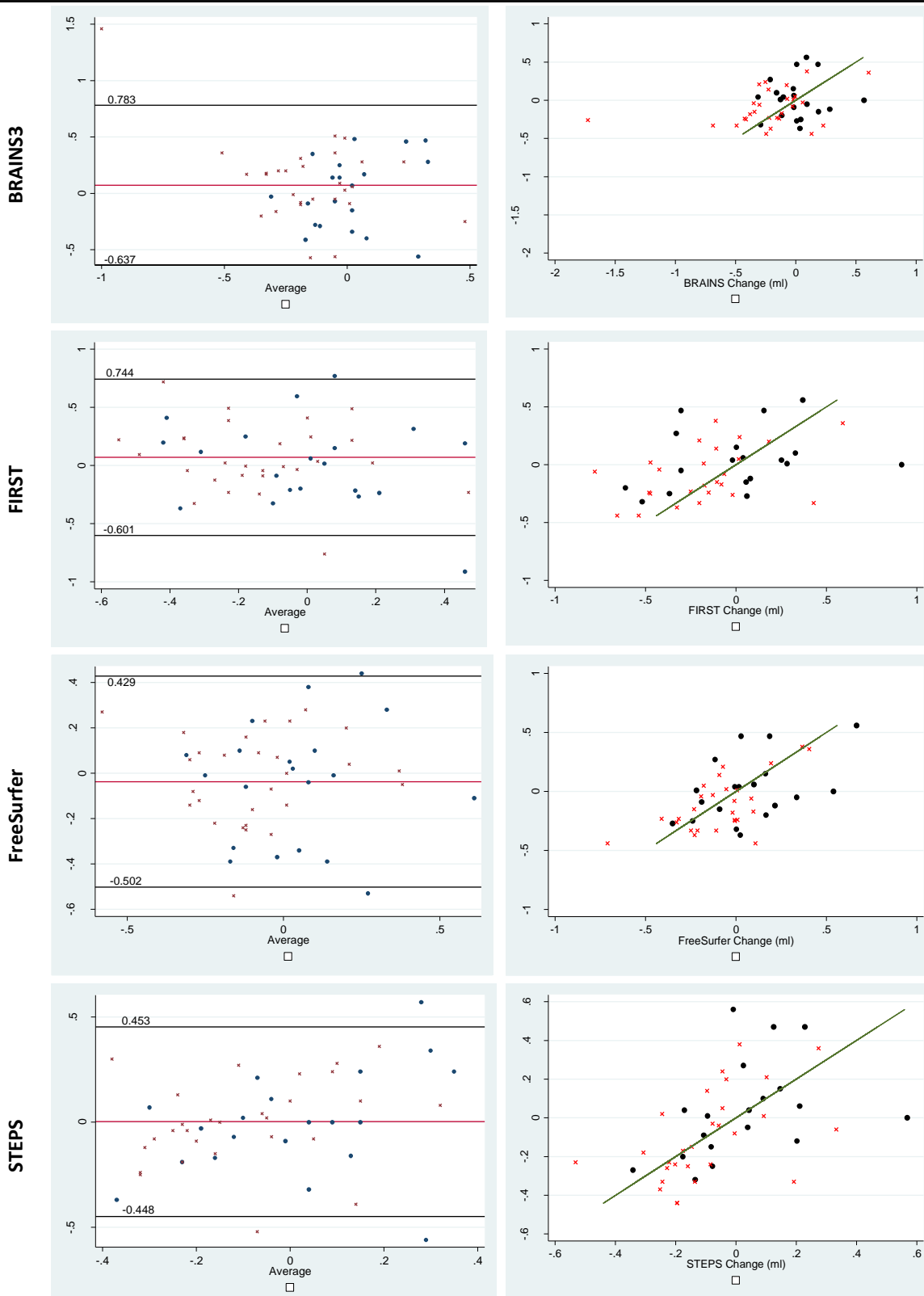


Figure 7-6. Bland Altman (left) and scatter plot (right) comparisons of automated longitudinal putamen volume change estimates (ml) against the manual gold-standard. Scatters of control (●) and HD (x) data-points between the manual volume change estimate (ml; y-axis) and the automated estimates of change (x-axis) also show the line-of-equality.

Estimates of percentage change in putamen volume over time are shown in Table 7-5; these range from -0.59 to -0.79% in the HD group over six months and from -1.22 to -3.91% over 15 months. All methods find significant between-group differences in putamen atrophy rate over 15 months but only FIRST and FreeSurfer (and manual with borderline significance; $p=0.057$) find this over 6 months. FIRST detected the strongest effect sizes over both 6- and 15-month intervals.

Table 7-5. 6- and 15-month percentage change in putamen volumes, with adjusted between-group differences, p-values and effect sizes.										
	6-Month % Change					15-Month % Change				
	Controls Mean (SD)	HD Mean (SD)	Adj. Between- Group Diff. (95% CI)	P- Value	Effect Size (95% CI)	Controls Mean (SD)	HD Mean (SD)	Adj. Between- Group Diff. (95% CI)	P- Value	Effect Size (95% CI)
Manual	.957 (2.132)	-.593 (4.616)	-1.801 (-3.653, .051)	0.057	-.450 (-1.226, .339)	.155 (3.387)	-2.080 (5.293)	-2.592 (-4.749, -.434)	0.019	-.642 (-1.451, .247)
BRAINS3	.515 (2.230)	-.750 (8.813)	-1.060 (-3.516, 1.396)	0.398	-.186 (-.691, .594)	.155 (2.706)	-3.911 (6.655)	-3.462 (-5.922, -1.002)	0.006	-.574 (-1.960, .060)
FIRST	1.127 (3.264)	-.786 (3.784)	-2.174 (-4.341, -.006)	0.049	-.594 (-1.320, .317)	.251 (4.089)	-2.451 (4.741)	-2.958 (-5.307, -.609)	0.014	-.756 (-1.653, .111)
FreeSurfer	.507 (1.951)	-.725 (2.107)	-1.110 (-2.207, -.013)	0.047	-.584 (-1.412, .410)	.654 (2.669)	-1.217 (3.282)	-1.492 (-2.975, -.010)	0.049	-.573 (-1.266, .223)
STEPS	.591 (2.697)	-.665 (2.437)	-1.067 (-2.346, .212)	0.102	-.524 (-1.195, .288)	.481 (3.083)	-2.077 (4.254)	-2.064 (-3.837, -.292)	0.022	-.587 (-1.361, .216)
Mean percentage change estimates are adjusted for scan interval. Adjusted between-group differences and effect sizes were modelled using generalised least squares regression controlling for age, gender, scan interval and study site. Effect sizes of between-group differences are reported with 95% bootstrapped CIs.										

Caudate Volumetry Method Comparison

Of the caudate segmentations the following were missing:

- BRAINS: one HD participant's caudate segmentation showed severe errors and therefore failed QC. This resulted in a missing 15-month change estimate.
- FIRST: segmentations from all three serial scans from three HD participants failed QC. Additionally, one control and one HD participant failed to output segmentations at six months. This resulted in one control and four HD participants with missing 6-month change estimates; three HD participants were missing 15-month change estimates.
- FreeSurfer: one control and three HD participants were missing 6- and 15-month change estimates due to QC failure on all three serial scans.
- STEPS: one control was missing a 15-month visit volume estimate.

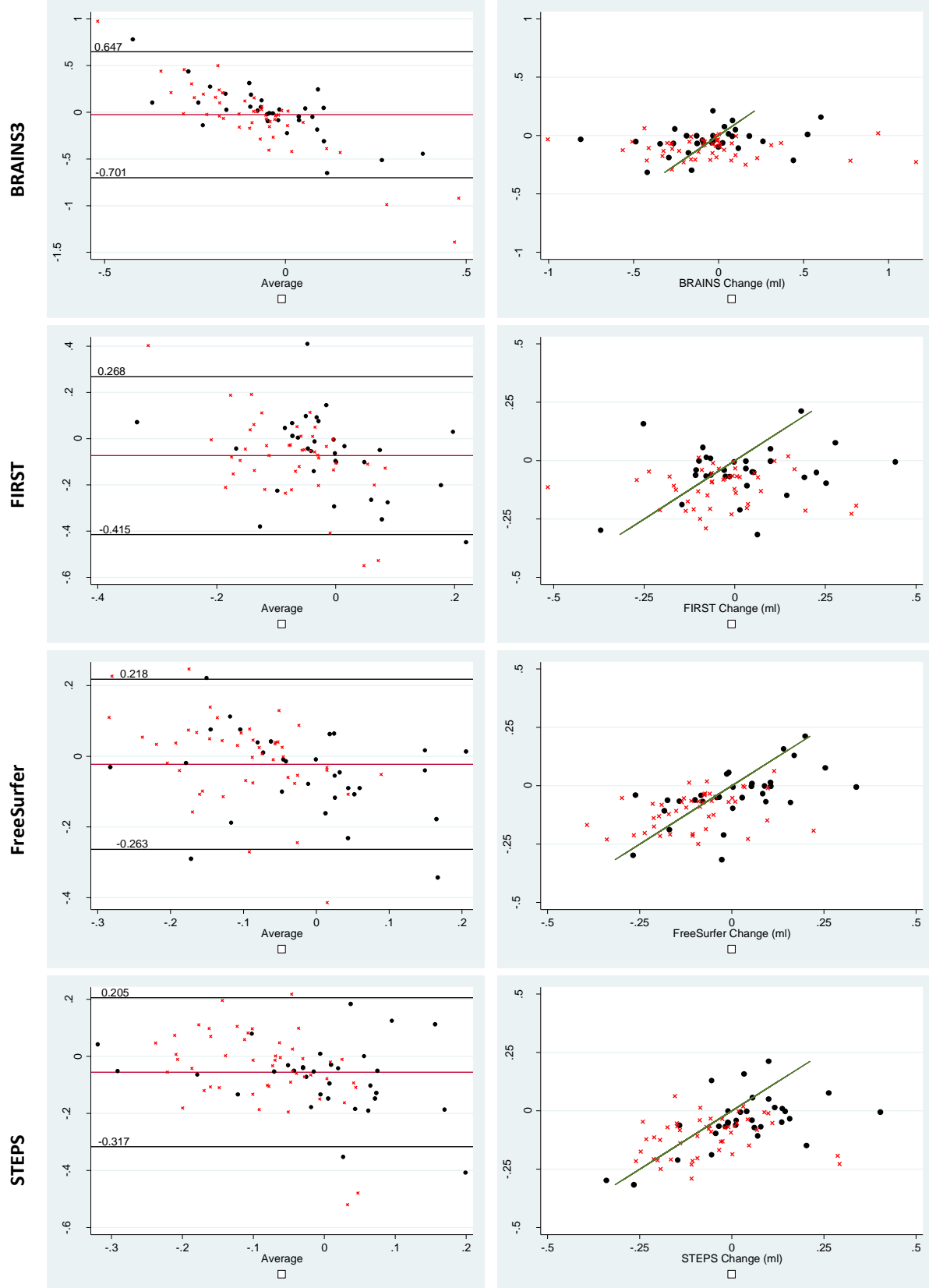
The CBSI was the only method to show a rate of change significantly different to zero in the control group over 6- (-0.042ml; $p=0.0441$) and 15-month (-0.083ml; $p=0.0003$) intervals. Significant atrophy over six months in the HD group (compared to zero) was detected by the CBSI (-0.107ml; $p<0.0001$), FreeSurfer (-0.098ml; $p<0.0001$) and STEPS (-0.065ml; $p=0.0007$). All methods reported significant change over 15 months in the HD group (compared to zero).

Table 7-6. Summary statistics of raw caudate 6- and 15-month volume change data.						
Method	Group	n	Mean (SD), ml	Mean Diff. (95% CI), ml (CBSI – Automated)	Ratio of Variance (95% CI) (CBSI /Automated)	Pearson's Rho (CBSI vs Automated)
6-Month Change						
CBSI	Controls	32	-.042 (.113)			
	HD	48	-.107 (.081)			
BRAINS3	Controls	32	-.047 (.282)	.005 (-.093 , .103) p=0.918	0.399 (0.282 , 0.566) p<0.001	0.2982
	HD	48	-.059 (.357)	-.049 (-.158 , .060) p=0.373	0.227 (0.170 , 0.304) p<0.001	-0.1265
FIRST	Controls	31	.021 (.161)	-.069 (-.133 , -.004) p=0.039	0.681 (0.473 , 0.981) p=0.039	0.1981
	HD	44	-.036 (.146)	-.076 (-.128 , -.025) p=0.004	0.542 (0.399 , 0.736) p<0.001	-0.0291
FreeSurfer	Controls	31	.009 (.144)	-.047 (-.091 , -.003) p=0.036	0.783 (0.578 , 1.059) p=0.110	0.5882
	HD	45	-.098 (.120)	-.005 (-.041 , .030) p=0.758	0.656 (0.494 , 0.872) p=0.004	0.3468
STEPS	Controls	32	.034 (.141)	-.075 (-.119 , -.031) p=0.001	0.799 (0.589 , 1.083) p=0.145	0.5584
	HD	48	-.065 (.123)	-.043 (-.082 , -.003) p=0.034	0.660 (0.494 , 0.880) p=0.005	0.1672
15-Month Change						
CBSI	Controls	32	-.083 (.115)			
	HD	48	-.191 (.130)			
BRAINS3	Controls	32	-.079 (.301)	-.004 (-.098 , .090) p=0.926	0.382 (0.279 , 0.523) p<0.001	0.5153
	HD	47	-.176 (.226)	-.012 (-.087 , .063) p=0.744	0.575 (0.428 , 0.773) p<0.001	0.0501
FIRST	Controls	32	-.015 (.157)	-.068 (-.133 , -.003) p=0.041	0.731 (0.510 , 1.049) p=0.088	0.1522
	HD	45	-.103 (.166)	-.093 (-.152 , -.034) p=0.003	0.786 (0.583 , 1.061) p=0.115	0.1462
FreeSurfer	Controls	31	-.037 (.162)	-.041 (-.091 , .009) p=0.104	0.699 (0.513 , 0.953) p=0.024	0.5596
	HD	45	-.228 (.200)	.037 (-.027 , .102) p=0.252	0.662 (0.492 to 0.890) p=0.007	0.2157
STEPS	Controls	31	-.017 (.168)	-.066 (-.124 , -.008) p= 0.028	0.696 (0.497 , 0.975) p=0.035	0.4255
	HD	48	-.145 (.153)	-.045 (-.094 , .004) p=0.069	0.847 (0.641 , 1.120) p=0.241	0.3037
Mean 6- and 15-month change and the SD of these estimates, outputted by each method. Comparisons with the CBSI included: paired t-tests to assess the significance of the between-method difference; two-sample variance-comparison tests; and pairwise correlations.						

FIRST, FreeSurfer and STEPS change estimates were significantly different from those of the gold-standard CBSI (Table 7-6). Variance was also higher in the majority of the automated estimates. This was particularly significant for the 6-month BRAINS estimates which ranged from -1.006ml to 1.162ml whereas the CBSI estimates ranged from -0.317ml to 0.212ml. FreeSurfer correlated most closely with the CBSI over both intervals (Rho 0.368-0.624), followed by STEPS (Rho 0.323-0.470).

Scatters and Bland Altman plots of the associations between these fully automated methods and the CBSI are shown in Figure 7-7. Multiple outliers in the BRAINS3 data have skewed the Bland Altman and scatter plots. These outliers are most severe in the HD data. No other clear estimate biases are apparent.

6-Month Caudate Volume Change Estimates versus CBSI Gold-Standard



15-Month Caudate Volume Change Estimates versus CBSI Gold-Standard

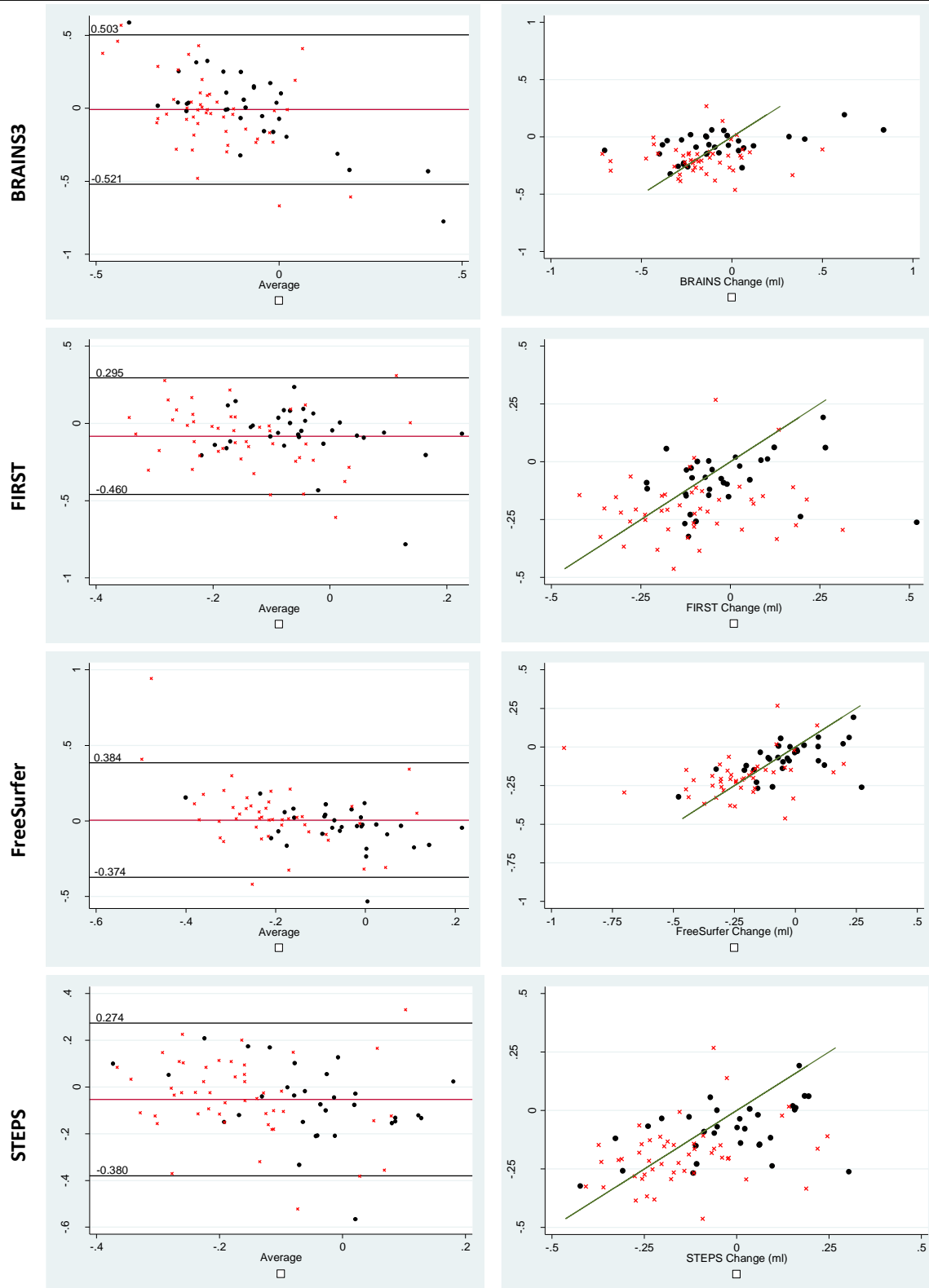


Figure 7-7. Bland Altman and scatter plot comparisons of automated longitudinal putamen volume change estimates (ml) against the CBSI gold-standard. Scatters of control (●) and HD (x) data-points between the CBSI volume change estimate (ml; y-axis) and the automated estimates of change (x-axis) also show the line-of-equality.

The estimates of percentage change in caudate volume are shown in Table 7-7. Only BRAINS and FIRST failed to detect significant between-group differences over six months. All methods showed significant differences over 15 months, with the CBSI and FreeSurfer showing the strongest effect sizes.

	6-Month % Change					15-Month % Change				
	Controls	HD	Adj. Between-Group Diff.	P-Value	Effect Size	Controls	HD	Adj. Between-Group Diff.	P-Value	Effect Size
	Mean (SD)		(95% CI)		(95% CI)	Mean (SD)		(95% CI)		(95% CI)
CBSI	-.480 (1.315)	-1.964 (1.597)	-1.586 (-2.254, -.919)	<0.001	-1.004 (-1.444, -.458)	-1.037 (1.424)	-3.922 (2.857)	-2.770 (-3.662, -1.878)	<0.001	-1.100 (-1.561, -.502)
BRAINS3	-.549 (4.492)	-2.098 (13.085)	-1.512 (-5.450, 2.426)	0.452	-.126 (-.576, .290)	-1.264 (5.004)	-6.194 (10.670)	-4.875 (-8.223, -1.528)	0.004	-.518 (-1.907, -.016)
FIRST	.400 (2.338)	-.570 (2.932)	-.923 (-2.136, .291)	0.136	-.330 (-.866, .184)	-.133 (2.554)	-2.129 (3.706)	-1.565 (-2.880, -.250)	0.020	-.477 (-1.928, .057)
FreeSurfer	.170 (1.897)	-1.860 (2.631)	-2.177 (-3.166, -1.189)	<0.001	-.991 (-1.472, -.435)	-.466 (2.188)	-4.678 (4.207)	-4.203 (-5.526, -2.879)	<0.001	-1.221 (-1.623, -.695)
STEPS	.534 (1.904)	-1.180 (2.535)	-1.625 (-2.593, -.658)	0.001	-.740 (-1.270, -.145)	-.128 (2.373)	-3.178 (3.428)	-2.619 (-3.730, -1.508)	<0.001	-.995 (-1.472, -.487)
Mean percentage change estimates are adjusted for scan interval. Adjusted between-group differences and effect sizes were modelled using generalised least squares regression controlling for age, gender, scan interval and study site. Effect sizes of between-group differences are reported with 95% bootstrapped CIs.										

7.3.4 Discussion

Putamen

The putamen should theoretically be a strong biomarker candidate in HD but structural analysis is difficult as the border contrast on T1-weighted MR images is not well-defined. Evidence from HD suggests that current analysis methods are underperforming. In this method comparison four automated techniques (BRAINS3, FIRST, FreeSurfer and STEPS) were compared with a manual gold-standard developed as part of this thesis (Section 7.2). Variability between methods, coupled with a lower sensitivity to HD-related change over both six and 15 months, confirming current volumetric biomarkers of the putamen to be suboptimal and substantially weaker than volumetric biomarkers of the caudate.

Visual analysis of the automated segmentations highlighted several regions of anatomically inaccurate segmentation errors, with STEPS judged to be visually the most reliable. All automated methods except STEPS included significant numbers of segmentations which would be deemed to contain moderate to severe errors. This highlights the need to visually inspect outputs from automated pipelines. Exclusion based on QC thresholds may help to increase the quality of the data from some methods. Alternatively, subsequent manual edits to automated outputs may be a reasonable compromise between the capacity to analyse large datasets but also maintaining an acceptable level of quality in the segmentations.

Additionally Table 7-3 highlights the inconsistency in inferior cut-offs applied by the different automated methods (and the manual method – example shown in Figure 7-2). FIRST, FreeSurfer and BRAINS display notable inclusion of WM inferior to the putaminal GM. As discussed in Section 7.2.2 the manual method (and STEPS) applies an inferior cut-off for reproducibility reasons. Consequently these methods are not biologically accurate estimates of full putaminal volume but are hypothesised to be more reproducible than the other automated methods. It is hypothesized that this exclusion of inferior GM will not have a differential effect on delineations of controls and HD scans and it is accepted that some signal may be lost.

FreeSurfer and STEPS reported estimates of putamen atrophy over both six and 15 months notably closer to the manual gold-standard than FIRST and BRAINS3. There was a suggestion in the STEPS data however that the HD group change may be underestimated using this method, particularly over the shorter 6-month interval. It was however FIRST that detected the strongest between-group effect sizes. This is evidence that strong effect sizes can be derived from data with high variance in outputs and segmentation quality and these effects should be interpreted accordingly.

In conclusion, the difficulty in delineating the putamen is reflected in the results of this method comparison. There were significant differences between results outputted from different methods and a lack of sensitivity to HD-related atrophy. Based on the visual assessment and the statistical comparison to the manual gold-standard, STEPS is concluded to be currently the most reliable automated method for longitudinal analysis of putamen volume. However, either larger numbers or intervals longer than six months will be required to detect HD-related putamen atrophy with this or other methods.

Caudate

In the caudate method comparison the gold-standard CBSI detected strong significant between-group differences over just six months with an effect size of -1.004 (95% CI -1.428, -0.475). The CBSI effect size over 15 months however was not much larger (-1.100 (95% CI -1.537, -0.482)). Again FreeSurfer and STEPS volumes were most consistent with this gold-standard, in this case also detecting the second and third strongest between-group effect sizes over six months (-0.991 (95% CI -1.457, -0.416) and -0.740 (-1.270, -0.145) respectively) are very strong and comparable to the CBSI.

It should also be noted that the CBSI measure was subject to stringent QC, with only those participants with CBSIs over both 6- and 15-month intervals passing QC included in the analysis (88.9% of the cohort). All automated outputs, unless severely erroneous, were included in the comparison. If outputs from these automated measures were inspected and poor quality data removed in this way the effect sizes may be even closer to this gold-standard.

Overall, FreeSurfer and STEPS appear to be strong automated candidates for longitudinal caudate volumetry which would facilitate large-scale volumetric analysis in comparison to the semi-automated CBSI method. Alternatively these methods could replace manual baseline caudate segmentations for use with the CBSI. Further work however is needed to validate this as a robust option.

Limitations

Limitations of this study are acknowledged. The putamen method comparison cohort was limited in sample size (n=48; 144 scans). This was, by design, set at this number owing to the time requirement for manual delineation (each scan required approximately one hour's analysis). It should be noted therefore that with larger numbers in the caudate method comparison these results cannot be directly compared to the putamen method comparison due to differences in power, although a sub-analysis suggests that these results largely hold if the same sample is used for both method comparisons (subset analyses not reported). Additionally, due to the difficulty in delineation of the putamen, this gold-standard measure was not as strong as the caudate (CBSI) gold-standard. Comparisons to the putamen gold-standard should therefore be interpreted according to the caveats mentioned in Section 7.2; in terms of measurement error and potential disease-related bias identified in this protocol.

In conclusion, the putamen volumetric biomarkers did not perform as well as the caudate biomarkers over the time intervals tested. The method used, as well as the structure chosen for analysis, also had a large effect on the quality, reliability and sensitivity of the measures. STEPS and FreeSurfer were found to be promising fully-automated alternatives to the manual and semi-automated gold-standards for striatal volumetry in large datasets. Stringent visual QC and possibly manual edits to these automated segmentations would ensure that these measures are maintaining the high quality of the gold-standards.

8 Volumetric Analysis of the Cerebellum

8.1 Background

Unlike the striatum which has been extensively analysed in HD, the cerebellum has received limited attention despite signs of possible cerebellar dysfunction; including motor incoordination and impaired gait. Direct anatomical connections are known to exist between the basal ganglia and the cerebellum (for a review see Wu & Hallett (Wu and Hallett 2013)), along which abnormal activity could propagate with negative consequences (Bostan and Strick 2010). Although notably spared in comparison to other brain regions (Rosas et al. 2003) there is evidence from autopsy (Jeste et al. 1984; Rodda 1981; Rub et al. 2013) and neuroimaging studies (Fennema-Notestine et al. 2004; Gomez-Anson et al. 2009; Ruocco et al. 2006; Scharmuller et al. 2013; Tabrizi et al. 2011) that cerebellar abnormalities are present in HD. Given this evidence it is surprising that in adult-onset HD the cerebellum has not been investigated in more detail.

In order to conduct an investigation into potential clinical associations of cerebellar pathology in HD (Chapter 13), a robust protocol for volumetric analysis was required. This study aimed to:

- a) Develop an accurate and robust manual delineation protocol for the cerebellum in MIDAS software.
- b) Apply the BSI technique to these cerebellar delineations.
- c) Conduct a cerebellar volumetry method comparison between change metrics derived from manual delineation, the cerebellar BSI and two automated software methods (FIRST and FreeSurfer).

8.2 Cohort: TRACK-HD

Data from the TRACK-HD study cohort was used for this protocol development (Tabrizi et al. 2009). Participants were sampled from the full cohort of 366: 123 controls, 120 preHD individuals, and 123 individuals with early HD. These participants were scanned at four sites on two scanner types: Philips Achieva (Leiden and Vancouver); Siemens Tim Trio (London and Paris) – full acquisition parameters are included in Appendix Section 23.3.

8.3 Protocol Development

There are three main challenges specific to cerebellar volumetric analyses: selection of voxel intensity thresholds which take into account the complex structural morphology and PVEs; definition of the cerebellum-brainstem boundary; and geometric distortion.

8.3.1 Voxel-Intensity Thresholds

All T1-weighted scans underwent bias correction (details in Section 4.2.1) to normalise voxel intensities across the image before thresholds were set and tested. Typically, manual delineations use intensity

thresholds to automatically delineate a range of voxels (examples in Section 4.2.2). This rough structural outline is then manually edited based on pre-defined criteria. Threshold selection is therefore highly important, determining the overall inclusion criteria; how ‘tight’ or ‘loose’ the fit of the outline to the borders. The cerebellum is a particularly difficult structure in this respect due to its branching and folding cortical morphology. Additionally, with the intricate internal structure of the cerebellar arbor vitae, a very small change in the scan intensity can have a large effect on the outline and consequently the volume estimate; an example of which is shown in Figure 8-1a. The PVE refers to the phenomenon that a single voxel can contain multiple tissue types due to finite image resolution. The high degree of curvature on the cerebellar folia results in relatively severe PVEs here compared with other brain regions. The PVE around the cerebellum is illustrated in Figure 8-1b. Due to these challenging aspects of cerebellar delineation it is important that thresholds are tested on a large sample across multiple scanner types, in both healthy controls and HD patients to find the most robust thresholds possible.

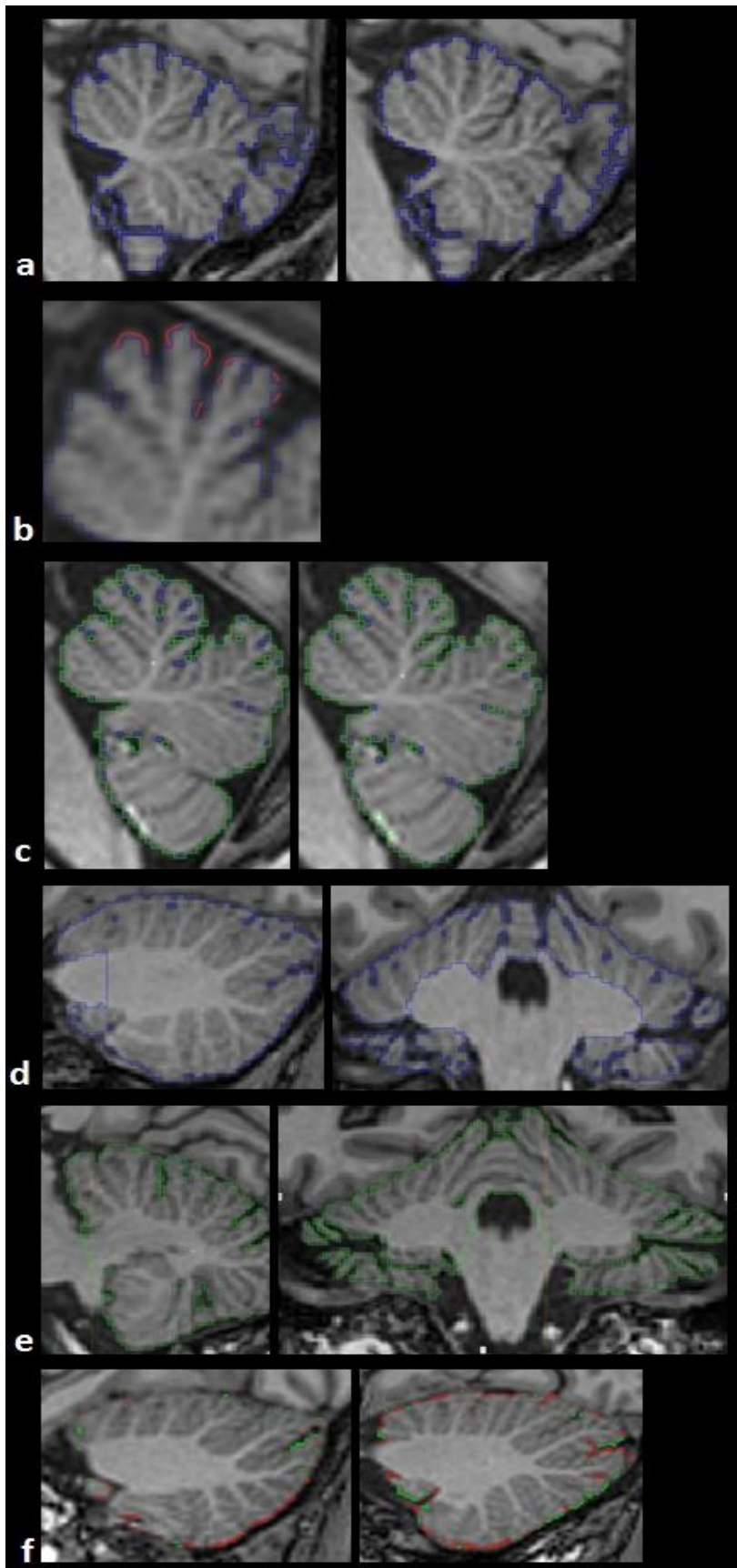


Figure 8-1. Cerebellum delineation protocol development examples:

a) Due to the branching structure of the cerebellum, small variations in voxel intensity can result in large delineation differences between two scans of the same participant.

b) Depiction of the PVE. The delineation (blue) is limited by the resolution of the scan (in this case 1mm³) and so is restricted in its accuracy on the highly curved borders of the cerebellum (highlighted in red).

c) The effect of increasing the lower voxel intensity threshold (as a percentage of mean brain intensity) on one cerebellar image slice: left) 60% (green) and 65% (blue); right) 65% (green) and 70% (blue).

d and e) Image slices from the two delineation protocols tested: d) Protocol 1; e) Protocol 2.

f) Visual representations of cerebellar BSI (red=contraction, green=expansion).

Initial investigations detected notable differences in the effect of lower threshold selection between scanner types (Philips and Siemens). Figure 8-1c shows an example of three lower thresholds (60%, 65% and 70% of mean brain intensity) which were tested on a cohort (n=20) of HD and control participants scanned on the two Philips Achieva scanners. For the majority of scans a lower threshold of 60% was found to be too low and to lose some of the finer detail of the cerebellar cortex. Increasing the lower threshold to 70% was too tight, cutting out regions of darker GM within the cerebellum. Several thresholds around 65% were tested but 65% was the final lower threshold selection. This was visually a good compromise between including all cerebellar tissue, limiting inclusion of CSF and capturing the complexity of the cortical outline. The same process was applied to Siemens scans (n=20) and a lower threshold of 70% was found to be optimal.

8.3.2 Boundary Definition

Initially, T1-weighted scans were rigidly registered to MNI305 atlas space. This enabled consistent application of landmark-derived cut-offs. There is a clear border along the superior edge of the cerebellum, separating it from the cerebrum. This can be manually delineated by eye in the coronal view. The complex morphology of the connection between the cerebellum and brainstem however results in difficulties defining the boundary here. This cannot simply be addressed by image intensity and therefore requires the definition of a reproducible cut-off. Two protocols were extensively tested (included in Appendix Section 23.5):

- Protocol 1) Anterior WM was removed in the coronal view up until the GM of the left and right cerebellar hemispheres merged at the vermis (Figure 8-1d)
- Protocol 2) The cut-off was defined in the sagittal view as a straight line from the most anterior superior GM to the most anterior inferior GM (Figure 8-1e).

The reproducibility of these protocols was tested with a week interval between analyses. The volume differences between repeated analyses using Protocol 1 were found to be <0.5% (n=4). Protocol 2 was found to be more easily reproducible and consistent over serial scans; <0.27% (n=4). Based on these results, protocol 2 was chosen as the preferred protocol for ongoing development.

8.3.3 Statistical Comparison of Thresholds

Background

With the development of a robust protocol (Protocol 2) based on visually-derived lower thresholds, a further analysis was conducted on a larger sample to assess the effect of this inconsistent threshold across scanner types (65% for Philips scans and 70% for Siemens) on outputted volumes.

Methods

The scans from a sample of 103 participants (52 control and 51 HD participants) from the TRACK-HD study, scanned on Philips (n=64; 34 controls and 30 HD participants) and Siemens scanners (n=39; 18 controls and 21 HD participants), were used for this investigation of intensity thresholds; demographics in Table 8-1. All scans were segmented twice each using lower thresholds of 65% and 70%. TIV was also measured using a previously published protocol (Whitwell et al. 2001).

Table 8-1. Cerebellum lower threshold comparison participant demographics.				
	Controls (n=52)		Early HD (n=51)	
Scanner	Philips	Siemens	Philips	Siemens
n	34	18	30	21
Gender: M/F	20/14	12/6	15/15	12/9
Age (years): Mean (SD)	45.9 (9.5)	47.0 (10.6)	47.1 (10.0)	49.7 (9.6)
Disease Burden ^a : Mean (SD)	NA	NA	392.8 (80.3)	370.9 (85.2)
^a Disease Burden as assessed by Penney et al. (Penney et al. 1997): (CAG-35.5) x age. NA= not applicable.				

Statistical Analysis

Paired t-tests were used to assess the significance of volume differences outputted by the different thresholds applied to the same scans. Regression models, adjusted for age, gender and TIV, were fitted to test the effect of lower threshold on cerebellar volume across scanner types (between different participant groups).

Results

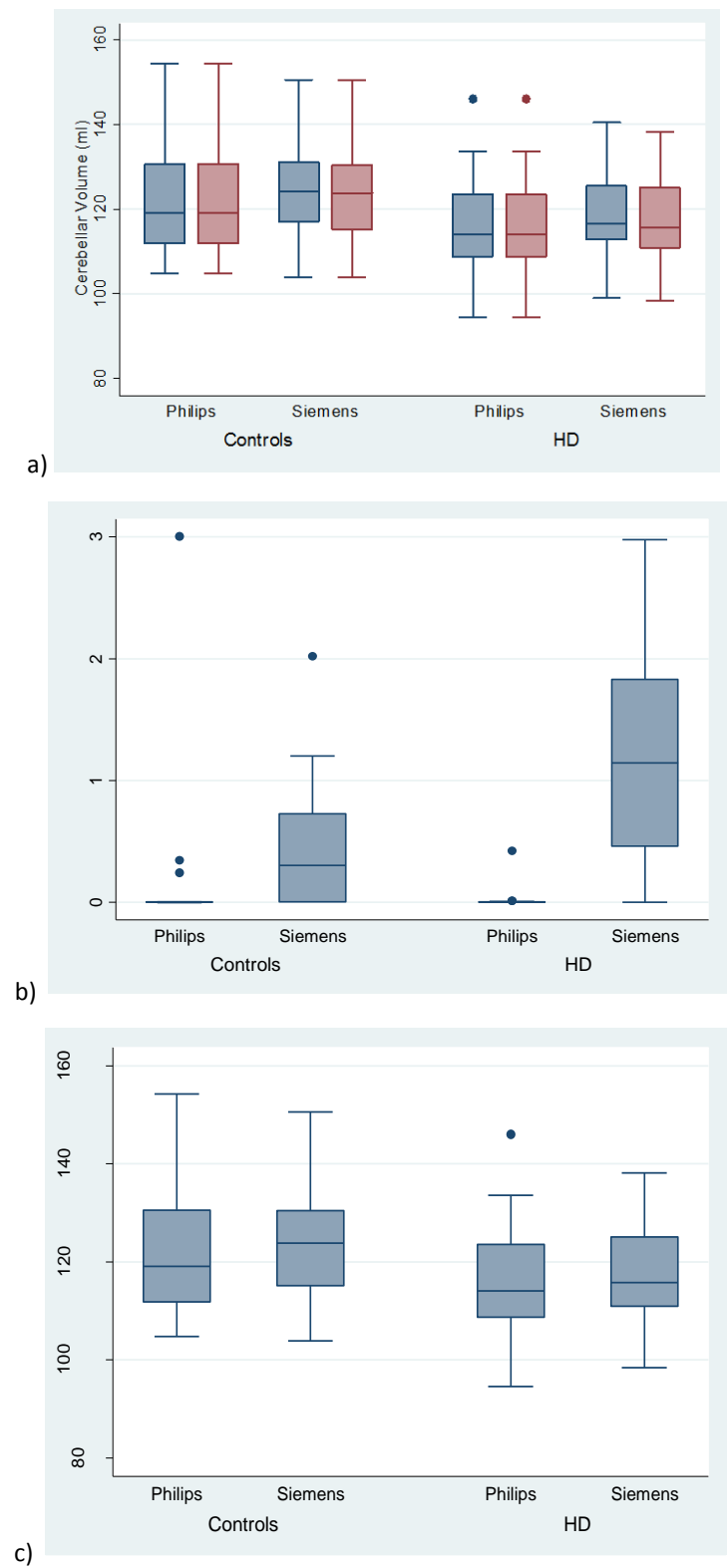


Figure 8-2. a) Unadjusted cerebellar volumes in control and HD groups scanned on Philips and Siemens scanners with lower thresholds of 65% (blue) and 70% (red) respectively; b) The differences in cerebellar volume outputted from 65% and 70% lower thresholds across groups and scanner types; c) Volumes outputted from both scanner types utilising the scanner-specific lower thresholds (65% for Philips and 70% for Siemens).

Figure 8-2a shows the cerebellar volumes outputted using both the 65% (blue) and 70% (red) lower thresholds. Combined across scanner types segmentations with a 65% threshold outputted significantly larger volumes than the 70% threshold segmentations in the control group (0.23ml (95% CI 0.08, 0.39) $p=0.0045$) and HD group (0.55ml (95% CI 0.10, 0.99) $p=0.0183$). These differences however were of very small magnitude. In the Philips scans this difference was present but not significant in controls (0.47ml (95% CI -0.07, 0.29) $p=0.2391$) but was significant in the HD group (0.14ml (95% CI 0.01, 0.26) $p=0.0401$). In the Siemens scans this difference was significant in both groups; controls (0.47ml (95% CI 0.18, 0.75) $p=0.0031$) and HD (1.32ml (95% CI 0.04, 2.59) $p=0.0433$).

With a 65% lower threshold volumes were examined across scanner types. There was no significant difference in control group volumes (3.50ml (95% CI -2.13, 9.14) $p=0.217$) but there was a highly significant difference in HD group volumes between scanner types (12.91ml (95% CI 6.79, 19.02) $p<0.001$). With a 70% lower threshold this between-scanner difference was, again, not significant in the control group (3.08ml (95% CI -2.52, 8.69) $p=0.274$) and was significant in the HD group (11.75ml (95% CI 5.35, 18.15) $p=0.001$), although slightly reduced.

When the lower threshold was increased from 65% to 70% the Siemens scan volumes were affected (decreased) significantly more than the Philips scans ($p<0.001$; Figure 8-2b). This effect was more pronounced in the HD group. These results are indicative of what is clear on visual inspection; the same lower threshold is 'looser', and consequently outputs larger volumes on Siemens scans than Philips scans.

When the scans were analysed with the scanner-specific lower thresholds (65% for Philips and 70% for Siemens; Figure 8-2c) there was no significant difference between the outputted cerebellar volumes between scanners in the control group (3.01ml (95% CI -2.63, 8.64) $p=0.288$) but there was a significant difference in the HD group (11.62ml (95% CI 5.24, 17.99) $p=0.001$).

Discussion

Visual inspection of thresholds for cerebellar segmentation initially suggested a 65% lower threshold for Philips scans and a 70% lower threshold for Siemens scans. This observation was supported by the finding of significantly smaller cerebellar volume estimates from Philips scans compared with Siemens scans when consistent lower thresholds were applied. With the visually-derived lower thresholds (65% for Philips and 70% for Siemens scans) this between-scanner difference remained significant in the HD group, although slightly reduced. These results suggest that, even with a higher lower threshold, Siemens scans produced higher volume estimates than the Philips scans.

This inconsistency between scanners is most likely due to subtle differences in tissue contrast between the cerebellar cortical GM and surrounding CSF, exacerbated by PVEs along this border. These between-

scanner effects are more extreme in the HD group, possibly due to atrophy-related increases in CSF spaces and consequently heightened PVEs. Alternatively this may be an artefact of the samples tested; the control group scanned on the Philips scanner showed a larger variance in volume estimates than the control group scanned on the Siemens scanner, possibly masking the effect found in the HD data.

Having two different groups scanned on each scanner type it is difficult to compare volumes directly. This is a limitation of the data available. The ideal test would be to scan the same cohort on both scanner types but for practical reasons this was not possible.

In conclusion between-scanner variability results in significant differences in cerebellar volumetry which cannot be completely removed by varying the protocol between different scanner types. Optimising the lower threshold selection for the scan acquisitions does however slightly reduce these differences and therefore study-specific tests are recommended to optimise intensity thresholds based on the data under investigation. The use of an inconsistent, adaptable protocol is acceptable for multi-site natural history studies. This variability however limits the potential of this region as a reliable biomarker in multi-site studies or trials with blinded image analysis.

8.3.4 Geometric Distortion

The cerebellum is located in the peripheral FOV and is consequently particularly susceptible to magnetic field inhomogeneity (Walker et al. 2014) potentially causing geometric distortions which could artificially increase or decrease the perceived and measured volume. The example in Figure 8-3A is of a participant whose serial scans show severe geometric distortion; there are notable distortions of the chin, neck and skull. A fluid registration from another participant between two scans (B) shows extreme cerebellar enlargement due to distortions of the cerebellar and brainstem region. All scan data should be inspected prior to cerebellum analysis for evidence of distortion and removed if it is judged that this distortion is sufficient to negatively impact on the accuracy of volumetric measures. Gradient warp correction based on phantom scanning and simulated data is also an option to reduce geometric distortion (Walker et al. 2014), although not implemented in this thesis.

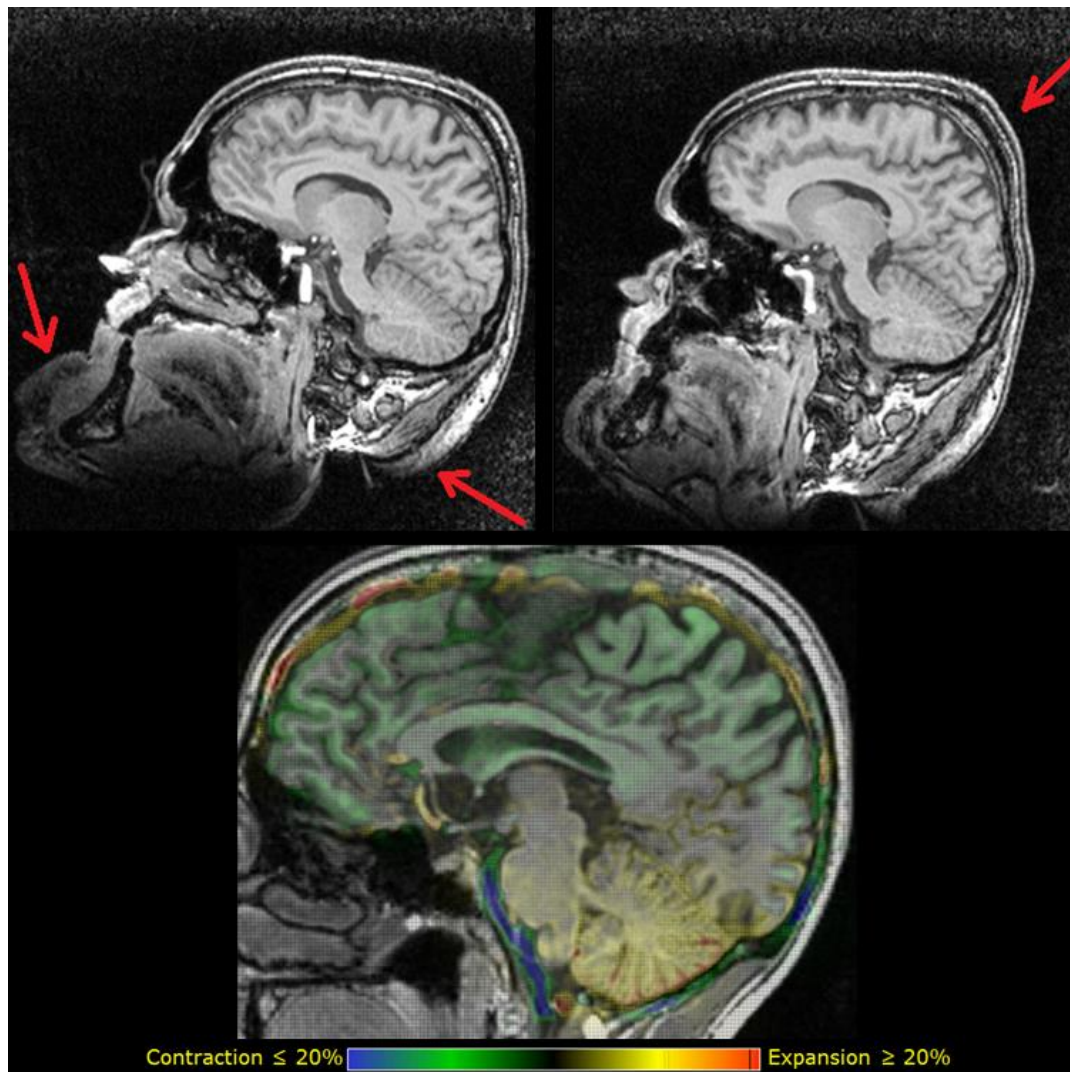


Figure 8-3. A) An example of a participant whose serial scans (left = baseline, right = follow-up) show severe geometric distortion most obviously affecting the chin, neck and skull (highlighted by red arrows). B) A fluid VCM between two scans showing extreme cerebellar enlargement due to distortions of the cerebellar and brainstem region.

8.4 Cerebellar Volumetry in HD: A Method Comparison

8.4.1 Background

In order to evaluate the reliability and strength of this novel manual delineation protocol for cerebellar volumetry it must be applied in an HD cohort and compared against alternative methods. This investigation will compare the manual method against several semi- and fully-automated alternatives with the aim of identifying the most reliable and sensitive method to detect between-group (HD versus control) differences in cerebellar volume and volume change over time. Since there is no validated gold-standard measure for cerebellar volumetry in HD, manual delineation is taken to be the gold-standard for the reasons outlined earlier. However, as little is published on manual delineation of the cerebellum these methods will be evaluated and compared based on the following criteria:

1. From autopsy evidence it is known that there is cerebellar atrophy in HD (Jeste et al. 1984; Rodda 1981; Rub et al. 2013) therefore do methods detect between-group differences cross-sectionally and longitudinally?
2. It is assumed that cerebellar atrophy will associate with the BBSI. How well do estimates of cerebellar volume change associate with the BBSI over the same interval?
3. It is hypothesized that reliable methods will be consistent with each other; with the caveat that they might be consistently showing bias. Between which methods are volumes in agreement/consistent?

The methods that were compared included manual delineation, FIRST, FreeSurfer and two newly developed methodological variants of a cerebellar KN-BSI. The BSI, described in detail in Section 4.2.11, is an automated direct measure of change which has successfully been applied to the whole-brain (Freeborough & Fox 1998), caudate (Hobbs et al. 2009) and ventricles (Freeborough & Fox 1997). The KN-BSI is an optimisation of this method for longitudinal, multi-site data (Leung et al. 2010). Direct measures of change are preferable to indirect measures (e.g. follow-up volume – baseline volume) as these theoretically reduce the potential for error, i.e. delineations at both time-points are required for indirect measures. Therefore the KN-BSI technique was applied to the cerebellum for inclusion in this method comparison; the customisations of the BSI for use on the cerebellum are described below.

8.4.2 Cerebellar KN-BSI Development

The cerebellum was manually delineated, using the protocol described in Section 8.3 and included in Appendix Section 23.5.2, on baseline and 24-month scans in MNI305 standard space. Serial scans were analysed in parallel (blinded to time-point) to ensure consistent application of cut-offs. The KN-BSI processing pipeline (detailed in Section 4.2.11) was customised for the cerebellum; this involved an initial whole-brain affine registration between serial scans, with differential bias correction, followed by a local rigid cerebellar registration (within the delineated regions dilated by two voxels). Traditionally the brain-brain registration is done within undilated brain delineations; for this customised pipeline these masks were dilated by eight voxels with the aim of creating a better estimate of the scaling factors between the scans by incorporating more scalp in the registration. Additionally, the two registrations are typically computed separately, resulting in the image being resampled twice; this pipeline was adjusted to conduct a single resampling step based on the composition of the two intermediate registrations. The aim here was to reduce resampling artefacts and consequently reduce blurring to improve sensitivity to change. K-means clustering was used to calculate the mean intensity of the GM, WM and CSF. Linear regression was then applied between these mean intensities and the results were applied to the scans to normalise the intensities. The region over which the BSI was computed was defined in two ways, which will be compared:

1. Incorporating information from just the baseline manual delineation; eroding and dilating this region by one voxel.
2. Utilising both baseline and follow-up delineations: an intersection region, common to both baseline and follow-up segmentations, was identified and eroded by one voxel; a union region, including voxels within segmentations from baseline, follow-up or both, was dilated by one voxel; the cerebellar BSI computation region was created from the dilated union and eroded intersect regions, i.e. subtracting the eroded region from the dilated region.

Intensity windows were based on the k-means derived results and the integral of change was calculated between the two overlaid scans within the BSI computation region between these scan-specific intensity windows. This k-means analysis was computed forwards and backwards and averaged to avoid potential bias. An example of the cerebellar KN-BSI overlay can be seen in Figure 8-1f.

8.4.3 Methods

Cohort

A sample cohort from the Leiden site of the TRACK-HD study (n=35; 19 controls and 16 HD participants), with baseline and 24-month scans, was used for this method comparison. Cohort demographics are detailed in Table 8-2.

Table 8-2. Cerebellum method comparison participant demographics.		
	Controls	Early HD
n	19	16
Gender: M/F	11/8	11/5
Age (years): Mean (SD)	48.0 (8.0)	46.7 (9.2)
CAG: Mean (SD)	NA	43.6 (2.6)
Disease Burden ^a : Mean (SD)	NA	357.8 (53.3)
^a Disease Burden as assessed by Penney et al. (Penney et al. 1997): (CAG-35.5) x age.		

Image Analysis

Cerebellum volume was analysed from baseline and 24-month scans using several methods - details of which are below. Three methods outputted volumes and indirect measures of change:

1. **Manual** delineations (protocol in Appendix Section 23.5.2).
2. Automated **FreeSurfer** software (version 5.1.0) run with the longitudinal pipeline (Reuter et al. 2012) - for an example FreeSurfer cerebellum segmentation see Figure 8-4a.
3. Automated **FIRST** analysis. Scans were initially registered to MNI305 standard-space using FMRIB's Linear Image Registration Tool (FLIRT (Jenkinson et al. 2002)). An additional local registration step was applied to optimise the alignment between cerebellar regions. Segmentations were computed,

followed by a boundary correction (using FMRIB's Automated Segmentation Tool, FAST (Zhang et al. 2001)) to prevent overlap between segmented regions, e.g. cerebellum and brainstem. An example of a FIRST cerebellar segmentation is shown in Figure 8-4b.

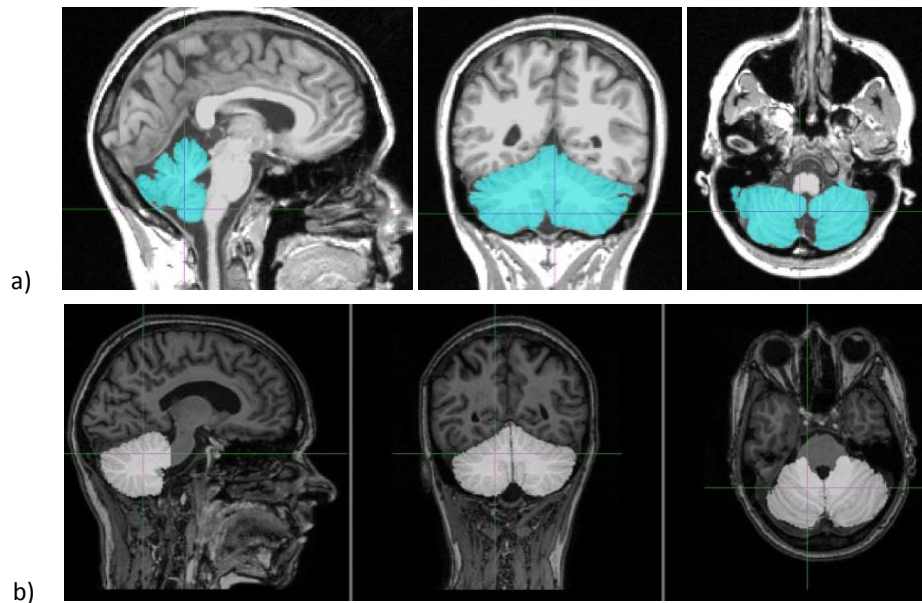


Figure 8-4. Examples of cerebellum segmentations from the two automated software methods in this method comparison:
a) FreeSurfer and b) FIRST.

Two variants of the KN-BSI provided direct measures of change:

1. **KN-BSI (baseline only)** – This method derived the integral using the baseline manual delineation only to define the computation area.
2. **KN-BSI (baseline + repeat)** – This method used both the baseline and follow-up delineations.

TIV was also outputted using the New Segment toolbox in SPM8 (validated against the semi-automated protocol of Whitwell et al. (Whitwell et al. 2001) for use in HD).

Statistical Analysis

Between-method differences in mean cerebellar volume estimates were assessed using paired t-tests. Generalised squares least regression models were fitted to assess between-group differences. These were adjusted for age, gender, TIV (cross-sectional analyses; to account for inter-subject variability in head-size) and scan interval (longitudinal only). Spearman's rank correlation coefficients were computed to assess agreement between methods. Correlation with the BBSI is also reported; here the BBSI is acting as a surrogate marker of general levels of brain atrophy.

8.4.4 Results

Cross-sectional cerebellar volumes extracted using manual delineation, FIRST and FreeSurfer are summarized in Table 8-3. Manual delineation and FreeSurfer reported relatively similar volumes and between-group differences. FIRST reported similar control group volumes but the volumes in the HD group were significantly higher than with manual delineation (12.0ml (95% CI 18.3, 5.8) $p=0.0010$) and FreeSurfer (5.3ml (7.0, 3.6) $p<0.0001$).

Table 8-3. Summary table of cross-sectional cerebellum volume measurements (ml), with adjusted between-group differences and p-values.

Methodology	Controls	HD	Adj. Group Difference	P-Value
	Mean (SD)	Mean (SD)	Mean (95% CI)	
Manual	121.214 (9.878)	111.855 (10.280)	-4.098 (-6.631, -1.565)	0.002
FIRST	124.388 (13.552)	123.881 (15.592)	.239 (-2.666, 3.144)	0.872
FreeSurfer	127.324 (10.949)	117.125 (11.393)	-4.374 (-7.245, -1.504)	0.003
Between-group differences were assessed using generalised least squares regression models adjusted for age, gender & TIV.				

Estimates of cerebellar volume change over 24 months are summarized in Table 8-4. Against expectations, analysis with FIRST reported volume expansion in both groups. Additionally, all other methods reported less cerebellar volume loss in the HD group compared with the control group; this effect was significant in the manual delineation data (-0.682ml; $p=0.037$). After adjustment for between-group differences in age, gender and scan interval this difference inverted to the expected direction in the FreeSurfer data.

Table 8-4. Summary table of longitudinal cerebellum volume change estimates over 24 months (ml; adjusted for interval), with adjusted between-group differences & p-values.

Method	Controls	HD	Adj. Group Difference	P-Value
	Mean (SD)	Mean (SD)	Mean (95% CI)	
Manual	-2.008 (2.339)	-.491 (1.965)	.682 (.041, 1.324)	0.037
FIRST	1.985 (4.865)	1.051 (3.455)	-.242 (-1.468, .983)	0.698
FreeSurfer	-1.311 (2.245)	-1.266 (1.891)	-.106 (-.725, .512)	0.736
KN-BSI (Baseline)	-.567 (1.054)	-.124 (.884)	.193 (-.100, .486)	0.196
KN-BSI (Baseline + repeat)	-.804 (1.240)	-.192 (.941)	.260 (-.060, .580)	0.111
Between-group differences were assessed using generalised least squares regression models adjusted for age, gender & scan interval.				

Unsurprisingly, the KN-BSI methods showed the highest pairwise correlation ($Rho>0.9$; Table 8-5). The manual method showed relatively similar 24-month change estimates to these KN-BSI estimates in both the control (0.875 and 0.753) and HD (0.835 and 0.879) groups. FIRST change estimates correlated very poorly with the BBSI, in the control group even showing a negative association. For all methods except FIRST the associations with the BBSI were higher in the HD group than the controls, possibly due to the higher signal in this group. FreeSurfer showed the closest correlation with the BBSI (0.75 in the HD group).

Table 8-5. Spearman's rank correlation coefficients for the associations between 24-month cerebellar change estimates (ml). Correlations with the 24-month BBSI (ml) are also included.

Method	Group	FIRST	FreeSurfer	KN-BSI	KN-BSI (Baseline+ Follow-up)	BBSI
Manual	Controls	-0.283	0.703	0.875	0.753	0.179
	HD	-0.100	0.579	0.835	0.879	0.588
FIRST	Controls		-0.156	-0.226	-0.381	-0.535
	HD		-0.147	0.115	0.088	0.006
FreeSurfer	Controls			0.740	0.787	0.319
	HD			0.494	0.550	0.750
KN-BSI (Baseline)	Controls				0.905	0.267
	HD				0.988	0.688
KN-BSI (Baseline + Repeat)	Controls					0.442
	HD					0.679

8.4.5 Discussion

A comparison of methods assessing cerebellum volumetry was conducted. As there is no gold-standard for this analysis in HD the methods were compared based on known disease effects and by assessing the correlation between methods and the BBSI – a surrogate marker for general levels of brain atrophy. Based on these criteria cross-sectional performance was strong in two of the three methods tested. Longitudinal estimates were weak; explanations of why this might be are discussed.

Manual delineation and FreeSurfer volume estimates performed strongly cross-sectionally, whilst FIRST seemed to lack sensitivity to disease-related volume reduction. FIRST analysis is based on a training set and the models cannot represent variations in shape and intensity that are not explicitly present in the training data leading to restrictions in permissible shapes. This is the most likely factor behind why FIRST is not optimal for assessing atrophied structures.

HD-related cerebellar atrophy is expected over 24 months in manifest HD patients based on previous neuroimaging (Fennema-Notestine et al. 2004; Ruocco et al. 2006) and post-mortem findings (Jeste et al. 1984; Rodda 1981; Rub et al. 2013). In the current study only FreeSurfer detected a higher rate of atrophy in the HD group compared with controls over 24 months and this difference was not statistically significant. The detection of cerebellar atrophy in the two previously published imaging studies, and lack of significant findings here, may be partially due to the HD cohorts studied; the other cohorts included HD patients with mid-stage disease and also patients with juvenile HD which is known to manifest with severe cerebellar atrophy (Sakazume et al. 2009).

There are several possible explanations for the lack of longitudinal sensitivity in these measures. Firstly, the manual method's sensitivity to subtle differences in voxel-intensity and tissue contrast (demonstrated

during protocol development – Section 8.3) may have reduced its sensitivity to cerebellar atrophy across serial images which may contain intensity fluctuations. FIRST, with its apparent insensitivity to cross-sectional disease-related volumetric effects, was unlikely to be able to detect longitudinal atrophy in this group. The BSI technique, shown to be a strong direct measure of change for the whole-brain, lateral ventricles and caudate, might not be appropriate or able to cope with the complex morphology of the cerebellar cortex (folia).

The increased rates of atrophy reported in the control group compared with the HD group (significant with manual delineations) could be due to differences in the cerebellar cortical border. In control subjects where atrophy is minimal the folia are tightly packed and therefore may be indistinguishable from neighbouring borders across a large part of the cerebral surface area. In pre- and manifest stages of HD atrophy is expected to have begun, potentially opening up CSF spaces between the folded cortical surfaces, facilitating boundary definition and consequently detection of atrophy. Alternatively there may be a contribution of geometric distortion artefacts: distortions caused by magnetic field inhomogeneity have been shown to increase with increasing distance from the isocentre of the scanner (Walker et al. 2014). Therefore it could be that the larger control brains, with cerebellums in the very peripheral areas of the FOV, were subject to more severe warping.

Future work could investigate whether gradient-warp correction improves results. Additionally, by performing region subtraction (A-B and B-A) between methods a qualitative assessment of the regions included in one protocol vs. another could be performed. This may help to localise delineation errors and inconsistencies. Overlap metrics such as Jaccard indices would also provide additional evidence for method correspondence. Alternatively, simulated atrophy experiments might allow testing of changes in volume where the answer is known and therefore give more support to the use of one technique over another.

In conclusion the manual delineation protocol developed in Section 8.3 and FreeSurfer software provide comparable options for cross-sectional between-group investigations of cerebellar volume in HD. None of the methods tested however were validated for use longitudinally, most likely due to the complexity and variability in the cerebellar cortical outline.

9 Evaluation and Application of FreeSurfer Cortical Thickness Analysis as a Biomarker in HD

9.1 Introduction

Cortical thinning occurs during the process of normal ageing (Jack et al. 1997) but post-mortem data shows this process to be abnormal and accelerated in neurodegenerative diseases such as HD (Halliday et al. 1998; Vonsattel & DiFiglia 1998). The development of advanced neuroimaging tools, most notably FreeSurfer software (details in Section 4.2.4 (Fischl & Dale 2000)), has enabled disease-related changes to the cortex to be measured *in vivo*. FreeSurfer software's cortical thickness analysis has been suggested as a biomarker in HD (Rosas et al. 2008). Results from its first use as an end-point in a clinical trial in HD were recently published as part of the PRECREST study (Rosas et al. 2014). There is however currently very little in the HD literature testing the suitability of this technique as a biomarker. Accordingly, this chapter aimed to assess FreeSurfer software's reliability, consistency and sensitivity as a biomarker in HD by conducting short exploratory methodological investigations (Section 9.4) and subsequently applying it to the PADDINGTON cohort (Section 9.5).

9.2 Image Analysis

The basic methodology of FreeSurfer software was described in Section 4.2.4. Details of the default processing pipeline and statistical analysis are given below:

All scans are initially run through the cross-sectional pipeline (Dale et al. 1999; Fischl et al. 1999). This produces outlines of the cortical border; the inner GM-WM boundary and the outer GM-CSF boundary. At this stage raw thickness estimates can be extracted from each region defined by the Desikan-Killiany Atlas (Desikan et al. 2006). These are calculated as the shortest distance between the GM-WM boundary and the GM-CSF boundary at every point across the cortex. For serial data-sets these analysed scans are fed into a longitudinal pipeline (Reuter et al. 2012). This creates a template for each participant in a common space between the serial scans which is thought to reduce bias created by asymmetrical registration of follow-up scans to baseline (Reuter et al. 2012; Reuter & Fischl 2011). All time-points are then registered to this space and processing is initialised with common information from the mid-space template. The temporal data within each participant is reduced to a single statistic at each vertex, e.g. mm change over the scanning interval, and smoothed. To generate magnitude and significance maps a study-specific template is created by averaging all participants' pooled data. This average is then smoothed and it is on this average template that the statistics are computed.

All outputs were visually inspected for quality. Only those that failed to produce a useable output were removed from analyses. No manual edits were made for any of the FreeSurfer analyses conducted in this thesis in order to report pragmatic, generalizable and reproducible results.

9.3 Statistical Analysis

General linear models are fitted to investigate between-group differences; results can be presented as magnitude, variance or statistical maps. Variance here is calculated as the square of the standard error of the between-group difference. It is widely recommended that statistical maps are corrected for multiple comparisons using FDR correction at $p < 0.05$ or lower (Genovese et al. 2002).

There is a choice of models in FreeSurfer software: a DODS ('different onset, different slope') model or a DOSS ('different onset, same slope') model. The DOSS model constrains the slopes of any continuous variables (e.g. age) to evolve at the same rate in all groups (disease group, gender and study site). In order to increase sensitivity to between-group differences the DOSS model is recommended over the DODS model (<http://freesurfer.net/fswiki/DodsDoss>). Figure 9-1 illustrates two modelled datasets: in 'A' age is related to cortical thickness in a consistent way between the four study groups; in 'B' age effects the groups in different ways. If the data looks like 'A' then the DOSS model is valid. If the data looks like 'B' the DODS model must be used (difference slopes between groups and genders as age increases).

Investigational regression analyses using the DODS model were conducted looking specifically at the effect of age on cortical thickness within each group, gender and study site. None of the resultant significance maps (with FDR correction) reported any significant differences in the effect of age on these variables, therefore it was deemed valid to use the DOSS model for all analyses.

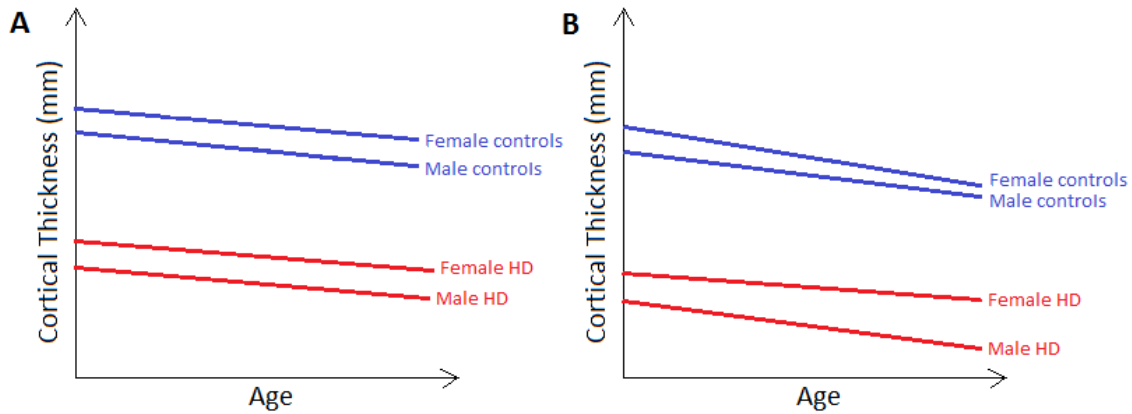


Figure 9-1. An illustration of two modelled datasets: in 'A' age is related to cortical thickness in a consistent way between the four study groups; in 'B' age effects the groups in different ways. If the data looks like 'A' then the DOSS model is valid. If the data looks like 'B' the DODS model must be used (difference slopes between groups and genders as age increases).

9.4 Methodological Investigations of FreeSurfer Cortical Thickness Software

Fully automated techniques are often thought of as being unbiased, objective and reproducible by different operators. This has been shown not to be the case with VBM analysis as varying the parameters can significantly affect the results produced (Henley et al. 2010). No such study has been published for FreeSurfer cortical thickness analysis. FreeSurfer offers multiple ways of assessing cortical thickness and change over time. There is no widely accepted best-practise and therefore variability in processing parameters may have an impact on the findings reported between studies. This section includes several short exploratory tests investigating the impact of commonly used methodological variations on the data.

9.4.1 A Comparison of Change Metrics

Background

Direct measures of cortical atrophy are calculated by FreeSurfer software prior to statistical comparisons. There are several options available based on rates and percentage change estimates. Rate of change is computed as the difference per time unit, so:

$$\text{Rate} = (\text{thickness 2} - \text{thickness 1}) / (\text{time 2} - \text{time 1}) = \text{mm/year}$$

Percentage change is the rate with respect to the thickness at the first time-point:

$$\text{Percentage change from baseline (PCB)} = \text{rate of change} / \text{thickness at baseline} \times 100$$

There is an additional option which aims to reduce noise by calculating the rate with respect to the average thickness between baseline and follow-up scans:

$$\text{Symmetrized percent change (SPC)} = \text{rate of change} / \text{average thickness} \times 100$$

SPC is the most highly recommended option by the FreeSurfer manual based on its assumed reduced level of noise and its symmetric nature, i.e. when reversing the order of time-point 1 and time-point 2 it switches sign; this is not true when comparing to baseline.

Based on simple calculations it is hypothesized that SPC will output larger estimates of change than PCB as the rate of change will be divided by a smaller (average) thickness. For example if the baseline thickness of a participant is 4mm and at one-year follow-up the thickness is 2mm, the following change metrics will be outputted:

- 1 PCB = $2 \div 4 \times 100 = 50\%$ per year
- 2 SPC = $2 \div 3 \times 100 = 66.7\%$ per year

Similarly, if there is a baseline difference in cortical thickness between groups, PCB estimates should produce larger between-group effect sizes than rate of change. For example, the same thickness reduction in mm will be reported as a higher percentage loss from participants with thinner baseline cortices, e.g. the HD group.

A short exploratory investigation was conducted to test these predictions and quantify the significance of these effects. Rate, PCB and SPC will be reported and compared to draw conclusions as to the robustness of these metrics for use in degenerative disease.

Methods

Cohort

PADDINGTON study participants were used for this metric comparison. 93 participants were scanned at baseline and 15 months. One of the 15-month visit scans (from an HD participant) failed initial scan QC. FreeSurfer analyses for three HD participants and one control failed QC leaving a cohort of 88: 36 controls and 52 HD participants. The demographic details of this cohort are shown in Table 9-1.

Table 9-1. Demographics of the PADDINGTON cohort with FreeSurfer analysed scans at both baseline and 15-month visits.		
	Controls	HD
n	36	52
Age; mean (SD)	52.0 (8.7)	47.7 (10.8)
Gender (M/F); n	16/20	19/33
Disease Burden ^a ; mean (SD)	NA	371.9 (89.3)
Site (1/2/3/4) ^b ; n	9/10/9/8	14/15/11/12
^a Disease burden calculated using the formula of Penney et al. (Penney et al. 1997). ^b Site 1=Leiden, 2=London, 3=Paris, 4=Ulm.		

Image Analysis

As described in Section 9.2, scans were run through the FreeSurfer (version 5.3.0) cross-sectional pipeline and then fed into the longitudinal pipeline which ran the analysis in a common space between scans. Default parameters were applied, with the recommended -3T flag to optimise analysis for 3T data.

Statistical Analysis

Regional thickness averages were outputted from all atlas regions and averaged within lobe. Rate of change (mm) between scans was calculated and adjusted for interval (rate/interval (years) x 1.25). Percentage change metrics were derived from these interval-adjusted rate metrics.

Between-group differences were computed using generalised least squares regression models, adjusted for age, gender and study site. Effect sizes were calculated as the estimated absolute adjusted mean difference of the metric between HD participants and controls, divided by the estimated residual SD of the HD group.

Paired t-tests were used to compare mean estimates outputted by the different metrics. Spearman's ranked correlation coefficients are reported between methods and a Bland Altman plot is used to represent the relationship between PCB and SPC estimates. Paired variance-comparison tests report between-method differences in estimate variability.

Results

All three metrics output the same pattern of atrophy rates across the cortex (Table 9-2). The highest and significant between-group differences were found in the occipital lobe, with results of borderline significance detected in the parietal lobe. In the frontal lobe there was a switch in the direction of the effect; controls showed more atrophy than the HD group.

Table 9-2. Rate and percentage estimates of cortical thickness change over 15 months (adjusted for interval) in averaged lobular regions, with adjusted between-group differences and effect sizes.

Lobe	Rate of Change				Percentage Change from Baseline				Symmetrized Percentage Change			
	Control mean (SD)	HD mean (SD)	Adj. Diff. (95% CI)	Effect Size (95% CI)	Control mean (SD)	HD mean (SD)	Adj. Diff. (95% CI)	Effect Size (95% CI)	Control mean (SD)	HD mean (SD)	Adj. Diff. (95% CI)	Effect Size (95% CI)
Frontal	-.035 (.078)	-.029 (.050)	.012 (-.016, .039) p=0.401	.237 (-.391, .899)	-1.332 (3.062)	-1.195 (2.058)	.378 (-.717, 1.472) p=0.499	.186 (-.376, .836)	-1.386 (3.116)	-1.222 (2.093)	.406 (-.708, 1.519) p=0.475	0.475 (-.364, .848)
Temporal	-.010 (.060)	-.027 (.066)	-.018 (-.044, .009) p=0.188	-.280 (-.769, .153)	-.315 (2.115)	-.933 (2.362)	-.657 (-1.593, .278) p=0.168	-.291 (-.718, .186)	-.336 (2.114)	-.964 (2.395)	-.669 (-1.611, .272) p=0.164	-.292 (-.709, .185)
Parietal	-.012 (.066)	-.037 (.057)	-.024 (-.059, .001) p=0.062	-.449 (-.936, .092)	-.521 (2.927)	-1.687 (2.651)	-1.119 (-2.256, .018) p=0.054	-.453 (-.900, .108)	-.564 (2.971)	-1.736 (2.723)	-1.122 (-2.280, .037) p=0.058	-.443 (-.882, .118)
Occipital	.001 (.058)	-.028 (.048)	-.025 (-.048, -.003) p=0.024	-.552 (-1.076, -.059)	.096 (3.116)	-1.526 (2.664)	-1.452 (-2.654, -.250) p=0.018	-.568 (-1.139, -.063)	.051 (3.089)	-1.572 (2.719)	-1.448 (-2.653, -.243) p=0.018	-.556 (-1.109, -.057)
Generalised least squares regression models were fitted to estimate the between-group differences in thickness change, adjusted for age, gender and study site. Effect sizes were calculated as the estimated absolute adjusted mean difference of the metric between HD participants and controls, divided by the estimated residual SD of the HD group. Dark grey highlights the significant (p<0.05) between-group differences and light grey highlights the borderline significant differences (p<0.08).												

As hypothesized SPC reported significantly ($p < 0.01$) larger change estimates than PCB in both groups across all regions, except control group occipital lobe estimates in which very little change was detected. PCB estimates also showed larger between-group effect sizes than rate in all regions except the frontal lobe.

As would be expected both percentage change metrics outputted estimates in identical ranked order. The rate of change metrics were highly correlated with this ranking; in the HD group Spearman's $Rho = 0.9991$, in the control group $Rho = 0.9982$. When PCB and SPC were directly compared there was an obvious bias in larger change estimates, with SPC overestimating this change (Figure 9-2). There was no significant difference in variance between these metrics in the control group ($p = 0.583$) but SPC showed significantly higher variance in the HD group ($p < 0.001$).

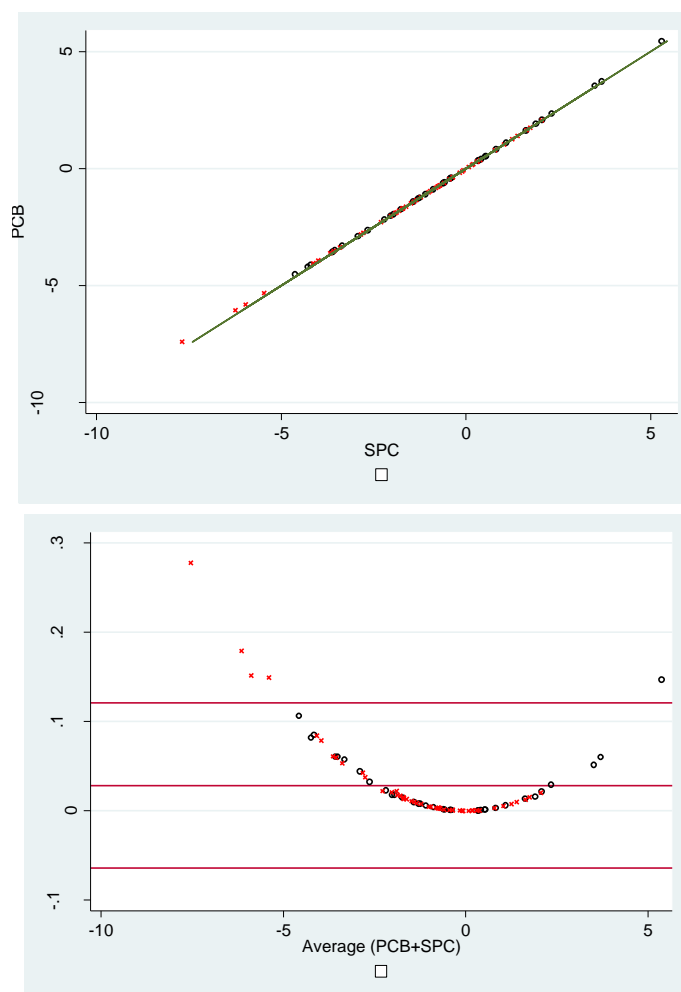


Figure 9-2. Scatter and Bland Altman plots comparing PCB and SPC estimates in whole-cortex thickness change.

Discussion

These results confirm initial hypotheses of data manipulation via choice of change metric. Estimates of PCB increased between-group effect sizes over rate estimates. SPC over-estimated change at higher estimates compared with PCB. These effects were found to be significant but not to change the overall pattern of results in this cohort.

In conclusion, rate of change and PCB are deemed to be the most robust of the three metrics tested. Rate of change is consistent across groups and easily interpretable but PCB may be more clinically relevant if it is believed that 1mm lost from an already thinned cortex will have more severe functional implications than 1mm lost from a non-atrophied cortex. There is a larger discussion in the HD research community as to whether absolute or percentage change metrics are more robust for reporting change, and consequently there is currently no consistent application between studies. The results of this investigation suggest that SPC should be used with caution as this metric artificially and significantly increases between-group differences.

9.4.2 The Effect of Increasing the Smoothing Kernel

Background

The smoothing of data is common in neuroimaging analyses in order to render the data more normally distributed and to correct for errors in the registration process (Ashburner & Friston 2000). Image data is convolved with a 3D Gaussian kernel so that voxel intensities become a weighted average of the surrounding voxels; the size of this kernel is user-defined. The general advice from the FreeSurfer community is that choice of smoothing kernel should depend on the sample size and the estimated size of the effect under investigation; larger samples and more localised atrophy require less smoothing. Kernels ranging from 5mm to 30mm FWHM are proposed by the software manual.

The *a priori* decision for the PADDINGTON data was to use a 10mm FWHM kernel (results reported in Section 9.5). This was based on the fact that 10mm was deemed to be a similar width to one gyrus thereby conserving localisation of effects. A published study assessing smoothing kernel variation on VBM data found that increasing the kernel diameter increased the extent of the findings (Henley 2010). It was therefore hypothesized that increasing the smoothing kernel for cortical thickness data would also increase the appearance of between-group differences by spreading the effects and potentially pooling multiple regions of significance into one larger region.

Methods

Cohort

All PADDINGTON participants with scans at baseline and 15 months were used for this investigation of smoothing kernel. Scans from three HD participants and one control failed QC leaving a cohort of 88: 36 controls and 52 HD participants. Cohort demographics were previously detailed in Table 9-1.

Image Analysis

To test the effect of increasing the kernel width, the between-group analyses assessing differences in thickness and rate of change were run using both 10mm and 20mm kernels; the *a priori* choice of the PADDINGTON study analysis and one with twice the diameter.

Statistical Analysis

Magnitude and significance maps of between-group differences in cortical thickness and rate of change over 6-, 9- and 15-month intervals, after analysis with both a 10mm and 20mm smoothing kernel, were outputted. These were qualitatively compared by eye to assess any differences in the extent or significance of findings.

Results

As hypothesized the larger 20mm smoothing kernel produced cross-sectional between-group difference maps with notably more significant and widespread findings (Figure 9-3). No significantly different between-group rates of change were detected over any interval with either size of smoothing kernel. The magnitude maps however showed larger clusters of differences with the wider kernel (Figure 9-4).

Discussion

Based on previously published findings from VBM data (Henley 2010), the choice of smoothing kernel was predicted to have a substantial effect on the extent of the significant findings on FreeSurfer cortical thickness maps. This was found to be the case, with the wider (20mm FWHM) smoothing kernel increasing the apparent significance and extent of the cross-sectional between-group differences and reducing the variability in the longitudinal change maps, thereby depicting a more succinct pattern of change.

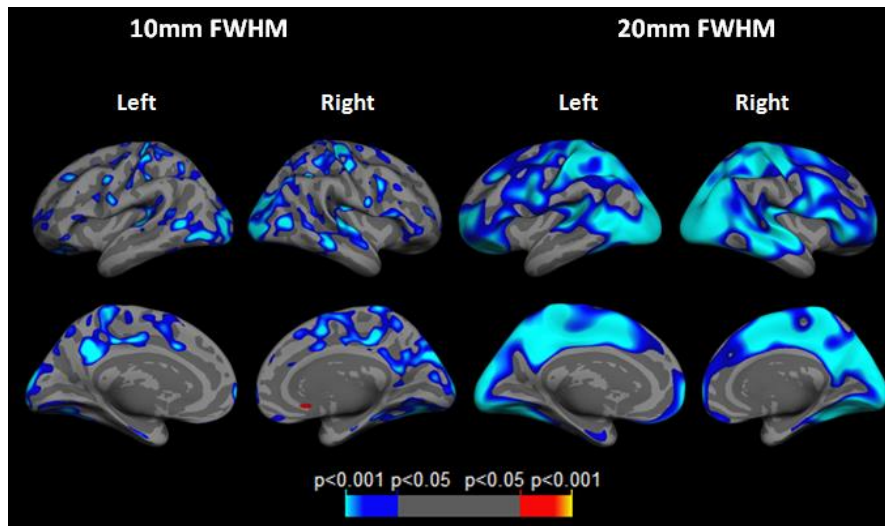


Figure 9-3. Significance maps of cross-sectional between-group differences in cortical thickness analysed with both a 10mm (left) and 20mm (right) smoothing kernel.

The wider smoothing kernel resulted in an artificial ‘spreading’ of the significant signal so much so that small, primarily occipital and parietal, regions of significance (as detected with the 10mm kernel) were expanded to include almost the entire medial cerebral mantle and approximately two thirds of the lateral cortex. This is most likely due to a ‘pooling’ of small regions of significance into one large area.

On the longitudinal change data the wider smoothing kernel had the effect of averaging conflicting regions of increased and decreased rates of atrophy between the groups, thereby reducing the spatial extent of between-group differences. Consequently these maps reported a more succinct pattern of change. Based on the patterns of the between-group differences on the magnitude maps presented, this seems to be an artificial simplification of noisy data. On maps where the signal is stronger the same effect as seen on the cross-sectional data would be expected.

In conclusion, the choice of smoothing kernel makes substantial differences to cortical thickness maps and therefore should be made prior to analysis based on the sample size and the estimated size of the effect under investigation. The kernel used should be clearly stated in all reports and justified according to these criteria.

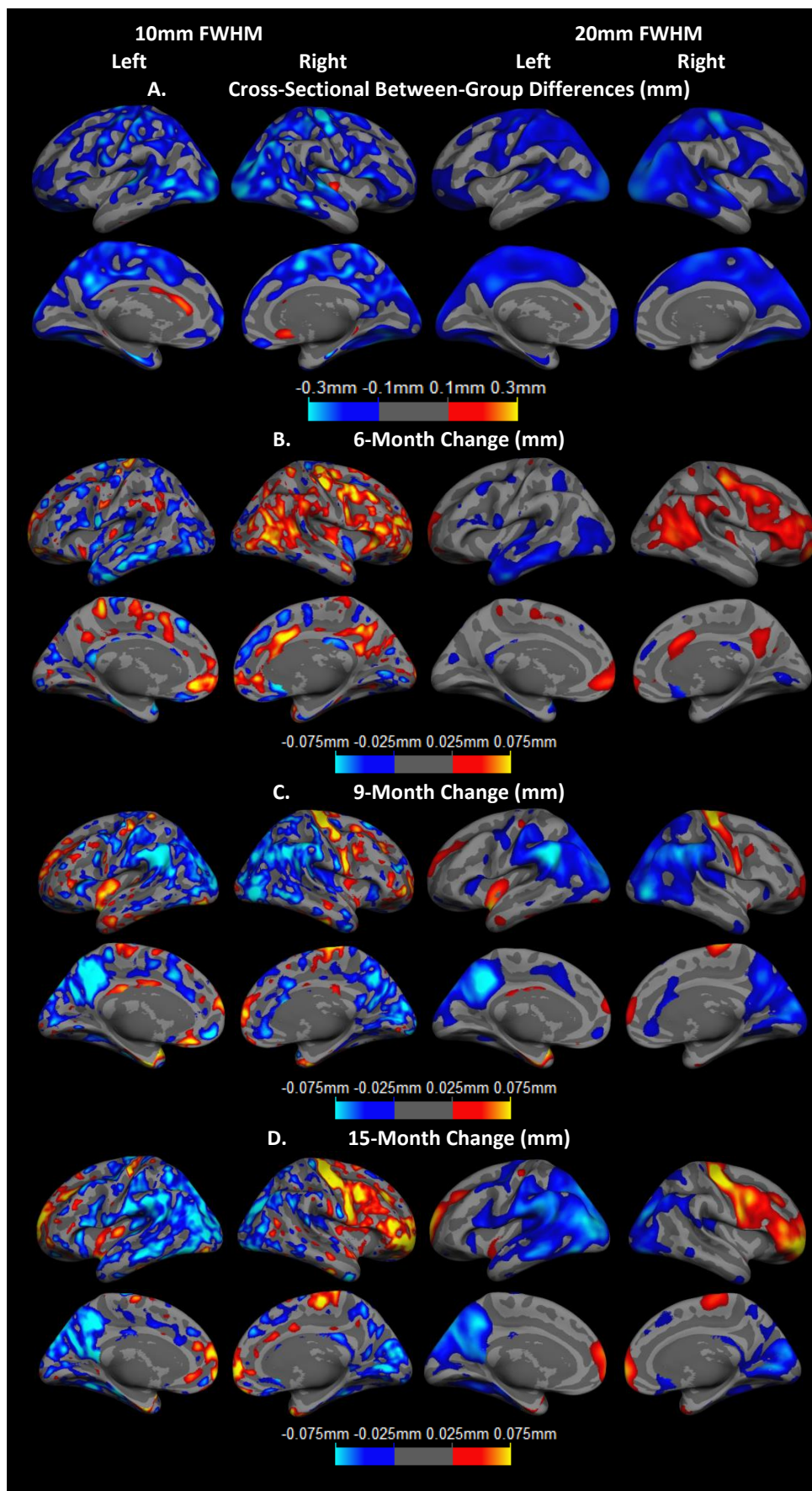


Figure 9-4. Magnitude maps of between-group differences in rate of change over 6-, 9- and 15-month intervals analysed with both a 10mm (left) and 20mm (right) smoothing kernel.

9.4.3 The Effect of Adjusting for Multiple Comparisons

Background

Cortical thickness significance maps are constructed from mass-univariate comparisons computed across 200,000 to 300,000 vertices. This large amount of statistical comparisons, each with a 5% chance of erroneously rejecting the null hypothesis, means that there are likely to be substantial amounts of false-positive results in unadjusted data. FreeSurfer provides the option to correct for multiple comparisons using FDR. This study aimed to conduct a qualitative evaluation of the extent to which FDR adjustment for multiple comparisons affected the sensitivity of FreeSurfer data to cortical thickness change over time in HD.

Methods

Cohort

All PADDINGTON study HD participants with serial scans were included in this analysis – see Table 3-1 for demographics.

Image Analysis

The FreeSurfer version 5.3.0 longitudinal pipeline was applied to the data following the details of Section 9.2. Unlike the other analyses in this chapter, the cortical thickness change metrics were used to compute change over time compared to zero, rather than detecting differences between groups.

Statistical Analysis

A repeated measures analysis of variance approach was used on the HD group data over six, nine and 15 months to assess cortical thickness change. The resultant statistical maps are reported with and without FDR correction for multiple comparisons at the threshold of 0.05.

Results

With no adjustment for multiple comparisons, significant change (both thickness increases and decreases) was seen in the HD group over just six months (Figure 9-5). Detectable atrophy was increased over nine months and widespread over 15 months. With adjustment for multiple comparisons significant atrophy was only evident over the 15-month interval.

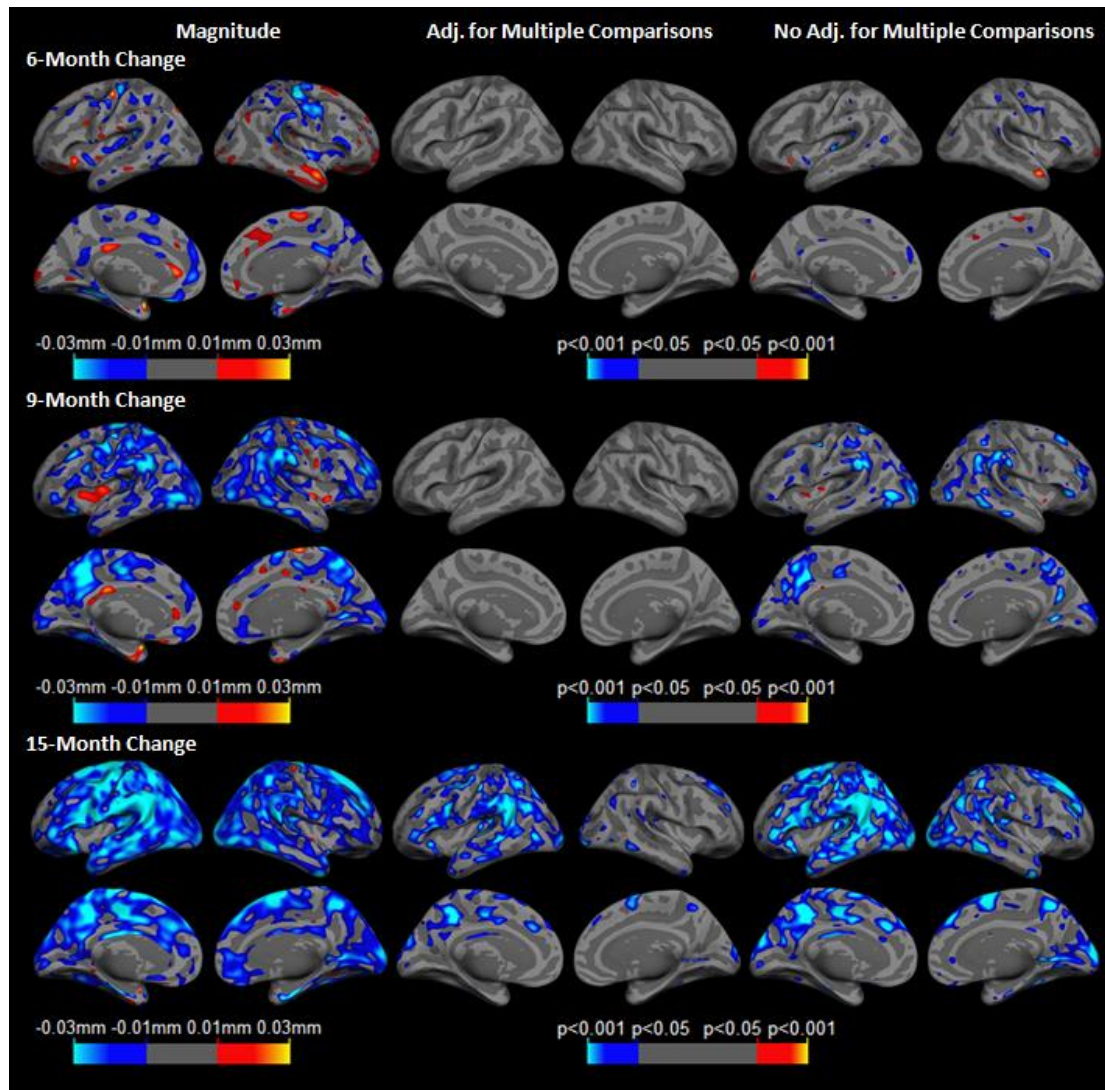


Figure 9-5. Magnitude (left) and significance maps, with (middle) and without (right) FDR adjustment (adj.) for multiple comparisons, of atrophy over six, nine and 15 months in HD participants - compared to zero.

Discussion

Adjustment for multiple comparisons made a substantial visual difference to the results detectable by FreeSurfer's cortical thickness software. Without adjustment for multiple comparisons significant change was detected in HD over just six months. The pattern of this change was however patchy and included regions of thickening, suggesting that this result contained a considerable amount of noise. The unadjusted results over nine months appeared more biologically plausible, with significant atrophy focused within parietal and occipital regions where it is known the majority of cortical thinning occurs in HD. Both adjusted and unadjusted data detected significant atrophy over 15 months. In conclusion, although more significant results can be obtained by not adjusting for multiple comparisons in cortical thickness analyses these results are not robust. All mass-univariate analyses should be corrected for multiple comparisons.

9.4.4 The Differences between FreeSurfer Versions (5.1.0 vs 5.3.0)

Background

It is widely assumed that later software versions are improvements on previous versions. A previously published study assessing differences between FreeSurfer versions 4.3.1, 4.5.0 and 5.0.0 found significant differences between version 5.0.0 and the two earlier versions (Gronenschild et al. 2012); $2.8\% \pm 1.3\%$ cortical thickness differences. The three most recent FreeSurfer software releases were versions 5.1.0, 5.2.0 and 5.3.0. Comparative differences in outputs from these software versions have not been published.

FreeSurfer version 5.1.0 was released in May 2011. This version was the first to introduce the longitudinal pipeline which uses a mid-space template between serial scans to avoid asymmetric processing bias. Version 5.2.0 was released in March 2013 but almost immediately retracted due to problems with pial and white surface creation affecting thickness and area measures. Version 5.3.0 was released with bug fixes in May 2013. This version added a '3T' flag to enable 3T-specific intensity correction and usage of a special 3T atlas for Talairach alignment and offers improved serial scan registrations (with the addition of cubic B-spline interpolation). In order to test the consistency and stability of FreeSurfer outputs between software versions, versions 5.1.0 and 5.3.0 were both run on the PADDINGTON study data.

Methods

Cohort

All PADDINGTON study participants with scan data at all three time-points were included in this study – see Table 3-1 for demographic details.

Image Analysis

All scans were run through the cross-sectional and longitudinal pipelines in FreeSurfer versions 5.1.0 and 5.3.0. Average thickness values from baseline and 15-month scans were extracted from all regions of the Desikan-Killiany Atlas (Desikan et al. 2006) and averaged for each lobe. Rates of change were computed and adjusted for interval $((V2-V1)/\text{interval} \times 1.25)$.

Version 5.3.0 data from four participants (one control and three HD participants) failed QC. Version 5.1.0 thickness data from six participants failed QC (three controls and three HD participants; three of whom (one control and two HD participants) also failed version 5.3.0 processing). This left a sample of 33 controls and 44 HD participants with longitudinal FreeSurfer data from both versions for use in this comparison.

Statistical Analysis

T-tests were conducted to assess the significance of mean differences in cortical thickness estimates between software versions and change from zero. Adjusted between-group differences and effect sizes

were modelled using generalised least squares regression controlling for age, gender, scanning interval and study site. Effect sizes of between-group differences are reported with 95% bootstrapped CIs.

Results

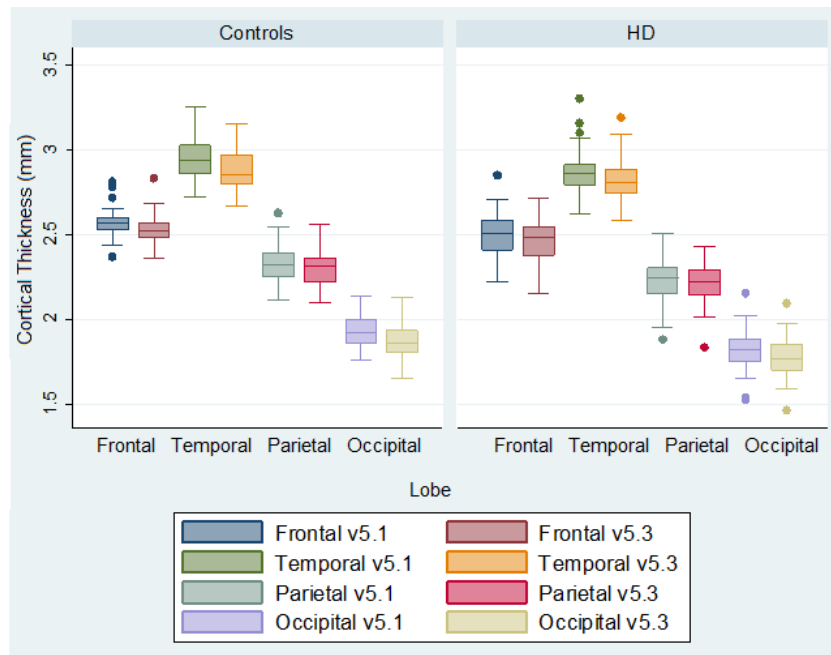


Figure 9-6. Cortical thickness averages (mm) within each lobe at baseline as outputted by both FreeSurfer versions 5.1.0 and 5.3.0.

With FreeSurfer version 5.1.0 the average thickness of the cortex in the control and HD groups was 2.55mm (SD 0.09) and 2.48mm (0.11) respectively. With version 5.3.0 this was estimated to be 2.39mm (0.09) and 2.32mm (0.11) in both groups respectively. This trend for thinner thickness estimates from version 5.3.0 compared with 5.1.0 was significant in all lobes ($p < 0.0001$ in all lobes; Figure 9-6).

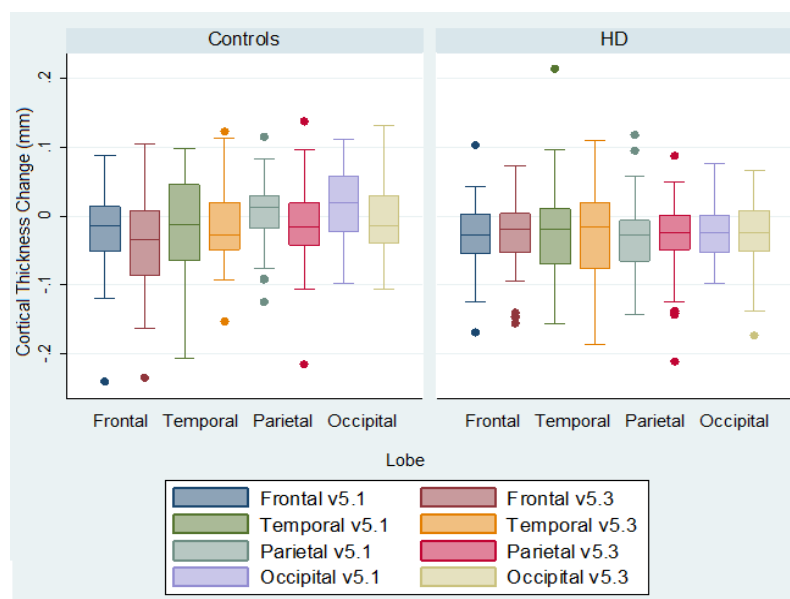


Figure 9-7. Rate of thickness change (mm) over 15 months within each lobe as outputted by both FreeSurfer versions 5.1.0 and 5.3.0.

In the control group, version 5.1.0 outputted slower rates of atrophy in frontal and temporal lobes compared with estimates from version 5.3.0 (Figure 9-7). In parietal and occipital lobes the version 5.1.0 mean control change estimates were positive (although not significantly greater than zero), indicating a non-significant trend towards cortical thickening. Version 5.3.0 indicated, on average, slight atrophy (again not significantly different from zero). The differences between these rates of change from the two software versions in the four lobes of the control group were of borderline significance ($p < 0.08$ in all lobes) in all except the temporal lobe. None of the rates in the HD data outputted by the two versions were significantly different. This difference in estimated control group atrophy had the expected result on the strength of between-group differences, with stronger effect sizes derived from the version 5.1.0 data, particularly in the occipital and parietal lobes (Table 9-3). These effect sizes are strong at -0.927 and -0.804 respectively. None of the between-group differences found by version 5.3.0 show statistical significance, although the between-group difference in the occipital lobe was of borderline significance ($p = 0.052$).

Table 9-3. Rate of lobular cortical thickness change over 15 months outputted by FreeSurfer versions 5.1.0 and 5.3.0, with adjusted between-group differences and effect sizes.										
Region	Version 5.1.0					Version 5.3.0				
	Controls	HD	Adj. Between-Group Diff.	P-value	Effect Size	Controls	HD	Adj. Between-Group Diff.	P-value	Effect Size
	Mean (SD)	Mean (SD)	95% CI		95% CI	Mean (SD)	Mean (SD)	95% CI		95% CI
Frontal	-.024 (.067)	-.024 (.067)	-.001 (-.026, .024)	0.937	-.022 (-.554, .628)	-.043 (.075)	-.027 (.052)	.019 (-.010, .047)	0.195	.357 (-.243, 1.001)
Temporal	-.012 (.066)	-.023 (.074)	-.015 (-.045, .014)	0.304	-.239 (-.738, .259)	-.016 (.057)	-.027 (.070)	-.013 (-.040, .015)	0.366	-.192 (-.663, .302)
Parietal	.003 (.053)	-.029 (.049)	-.035 (-.057, -.013)	0.002	-.804 (-1.413, -.183)	-.016 (.066)	-.036 (.060)	-.021 (-.048, .006)	0.130	-.374 (-.901, .163)
Occipital	.019 (.054)	-.021 (.045)	-.040 (-.062, -.018)	<0.001	-.927 (-1.523, -.314)	-.002 (.059)	-.026 (.050)	-.023 (-.047, .0002)	0.052	-.482 (-.968, .137)
Adjusted between-group differences and effect sizes were modelled using generalised least squares regression controlling for age, gender and study site. Effect sizes of between-group differences are reported with 95% bootstrapped CIs.										

Discussion

A previous study has shown significant differences in cortical thickness estimates outputted between different FreeSurfer versions (Gronenschild et al. 2012). No such investigation has looked into a potential effect of upgrade on the latest versions released by FreeSurfer (5.1.0 and 5.3.0). Both of these versions were run on the PADDINGTON data and significant differences were detected; version 5.1.0 reported thicker cortices and displayed greater sensitivity to between-group differences in rate of change over time than the newer version (5.3.0). With no gold-standard benchmark available it cannot be concluded that one or other version is superior but differences this large highlight a worrying lack of consistency. These results also have implications for the interpretation and comparability of previously published FreeSurfer cortical thickness findings.

The between-version difference in sensitivity to between-group thickness change was driven by reduced and negative (cortical thickening) atrophy rate estimates in the control group detected by version 5.1.0 but similar rates of atrophy estimated by both versions in the HD group. Cortical thickening, although potentially possible with neural plasticity (Maguire et al. 2000), is not biologically plausible at the group level and is therefore more likely an artefact of the image analysis.

It is not clear why these software versions should be outputting varying results from the control data but not the HD group data. Cortical atrophy is known to be present in early-stage HD, therefore one difference between the groups is likely to be increased spacing between the folds of the cortical gyri in the HD group compared with controls, potentially facilitating cortical thickness analysis. It could be that improved intensity normalisation and registration procedures in the newer version (5.3.0) enabled more accurate detection of control group atrophy, despite the close alignment of neighbouring gyri, thereby detecting real age-related change.

In conclusion the upgraded FreeSurfer software version (5.3.0) is not as sensitive to between-group differences in atrophy rates as the previous version (5.1.0). This may be a closer estimate to the biologically accurate rates of change but weakens this methodology as a biomarker candidate in HD. The continuing development of the software which results in such inconsistency in outputs is indicative of a methodology that is not yet ready for use on sensitive clinical trial data. Further optimisations and validations may however develop this into a robust and valuable biomarker of cortical atrophy.

9.5 Cortical Thinning as a Marker of Disease Progression in Early HD

9.5.1 Background

Six studies to-date have detected reduced cortical thickness in HD compared with control groups cross-sectionally, in far- to near-onset preHD (Nopoulos et al. 2010; Rosas et al. 2005; Tabrizi et al. 2009; Wolf et al.

2013), and early- to late-stage symptomatic HD (Rosas et al. 2002; Rosas et al. 2008; Tabrizi et al. 2009). One observational study reported longitudinal thinning of up to 8% over one year within HD patients (Rosas et al. 2011). No control group was included in these analyses however, making it difficult to identify disease-related longitudinal changes.

The first use of cortical thickness analysis in a clinical trial in HD was published in 2014 (Rosas et al. 2014); a phase II clinical trial of Creatine in preHD. A within-subject repeated measures analysis was conducted, quantifying the extent and localisation of cortical atrophy over six months in controls and preHD individuals either treated with Creatine or placebo. The preHD placebo group showed significant thinning of up to 5% per year ($p < 0.0001$) in certain regions. The preHD Creatine treated group showed less thinning. A direct statistical comparison across the whole-brain showed small areas of significant differences in these rates. When regional thickness change estimates were extracted from these preHD treatment and placebo groups, significant differences between rates of thinning were detected in regions including the precentral, occipital, superior frontal and supramarginal gyri. Numbers in this study were however small (22 placebo and 25 treated preHD participants) and results in healthy controls were not reported.

To fully evaluate cortical thickness as a biomarker of disease progression, several characteristics of the measure require investigation:

1. Variability in the normal population
2. Sensitivity to cross-sectional disease-related differences compared with healthy controls
3. Ability to detect significant changes over time compared with controls, ideally over short time intervals
4. Capacity for application across multiple sites (important for large studies and trials)
5. Reliability and consistency in findings of the natural progression of disease
6. Sample-size requirements for (realistic) hypothetical treatment effects.

This study aimed to evaluate cortical thickness as a biomarker based on the criteria listed above. This was the first study to assess the longitudinal sensitivity of cortical thickness measurements in a relatively large early-stage HD cohort compared with controls, over short and varying time intervals and across multiple scanners. Effect sizes of regional cortical change will also be evaluated in the context of other promising markers of disease progression in HD, as a way of comparing the relative sensitivity of the different measures. The results of this study will inform the decision of whether cortical thickness measures, and in what form, are included in a larger comparison of biomarkers in HD (Chapter 18).

9.5.2 Methods

Cohort

101 participants (61 early HD and 40 controls) were scanned at baseline as part of the PADDINGTON study. At six months 97 participants were scanned (57 early HD and 40 controls) and 93 were scanned at 15 months (56 early HD and 37 controls). Participant demographics are detailed in Section 3.1.

Image Acquisition

T1-weighted scans were collected from four 3T scanners: Philips Achieva (Leiden), Siemens Tim Trio (London), Siemens Verio (Paris) and a Siemens Allegra (Ulm). Acquisition parameters have been described previously (Section 3.4).

Image Analysis

All raw scans were assessed for quality; specifically motion and distortion artefacts and sufficient tissue contrast. Cortical thickness analysis was conducted using FreeSurfer software (version 5.3.0). Default parameters were used, along with a recommended -3T flag which optimises analysis for 3T data. All scans were initially run through the cross-sectional pipeline (Dale et al. 1999; Fischl et al. 1999), and subsequently fed into the longitudinal pipeline (Reuter et al. 2012). This created a template for each participant in a common space between the serial scans, thought to reduce bias created by asymmetrical registration of follow-up scans to baseline (Reuter et al. 2012; Reuter & Fischl 2011). All time-points were then registered to this space and processing was initialised with common information from the mid-space template. At each stage of the processing segmentations and registrations were visually inspected for accuracy.

Following the completion of all within-subject analysis, raw thickness measures were extracted from each region defined by the Desikan-Killiany Atlas in all scans (Desikan et al. 2006). To generate magnitude and significance maps for between-group comparisons a study-specific group template was created by averaging all participants' pooled data. This average was smoothed with a 10mm kernel; deemed to be optimal for removing noise whilst retaining localisation of signal. It is on this average template that the cross-sectional statistics were computed. For the longitudinal analyses the temporal data within each participant was reduced to a single statistic at each vertex and reported in mm change over the scanning interval (standardised to six, nine and 15 months exactly). These change metrics were smoothed with a 10mm kernel and between-group differences computed.

Statistical Analysis

General linear models were fitted to investigate between-group differences (adjusting for age, gender and study site) to produce magnitude, variance and statistical ($p < 0.001$ to $p < 0.05$) maps. Magnitude maps are reported as the between-group difference in mean thickness (mm) and thickness change (mm per interval).

Variance is calculated as the square of the standard error of the between-group difference. Statistical maps were corrected for multiple comparisons using FDR correction at $p < 0.05$ (Genovese et al. 2002).

The raw Desikan-Killiany atlas region thickness data was analysed separately. For longitudinal analyses the raw thickness measures were adjusted for interval. Generalised least squares regression models were fitted (adjusted for age, gender and study site) to assess between-group differences in thickness and change over six, nine and 15 months. Effect sizes are also reported; calculated as the estimated absolute adjusted mean difference of the metric between the HD and control groups, divided by the estimated residual SD of the HD group. Effect sizes have been inverted such that a positive effect size suggests between-group differences in the 'expected' direction, i.e. positive values represent greater thinning in the HD group compared with controls. An effect size of one implies that the mean change in the HD group is one SD higher than that in controls. These are reported with bias-corrected and accelerated bootstrapped 95% CIs based on 2000 replications (Carpenter & Bithell 2000).

9.5.3 Results

Four participants failed the longitudinal FreeSurfer pipeline (three HD participants and one control, all from the Ulm site (Siemens Allegra)). This left a total of 39 controls and 52 HD participants with 6-month change data, 36 controls and 47 HD participants with 9-month data, and 36 controls and 52 HD participants with 15-month data.

Cross-sectionally the average thickness across the whole cortex for the control group was 2.39mm (SD 0.09; 2.25 to 2.63mm) and 2.30mm in the HD group (SD 0.12; 2.05 to 2.55mm); the adjusted between-group difference was -0.11mm (95% CI -0.14, -0.07; $p < 0.001$). Widespread regions of reduced cortical thickness in the early-stage HD cohort compared with the control group can be seen (Figure 9-8A). In certain regions this is up to ~0.3mm thinner; the equivalent of a ~12.6% reduction from the control group average thickness. Despite significant cross-sectional findings no rates of cortical thinning in the HD group reached statistical significance in comparison to control rates of change over 6-, 9- or 15-month intervals - presented in Figure 9-8 B-D respectively. Trends in the magnitude maps suggest that HD-related atrophy may be more severe within the left hemisphere (although this was not directly tested). Several regions within the right hemisphere, notably including the precentral gyrus, showed (non-significant) slower rates of cortical thinning in the HD group compared with controls. It should be noted however that the precentral gyrus showed high variance levels in both the cross-sectional and longitudinal between-group analyses.

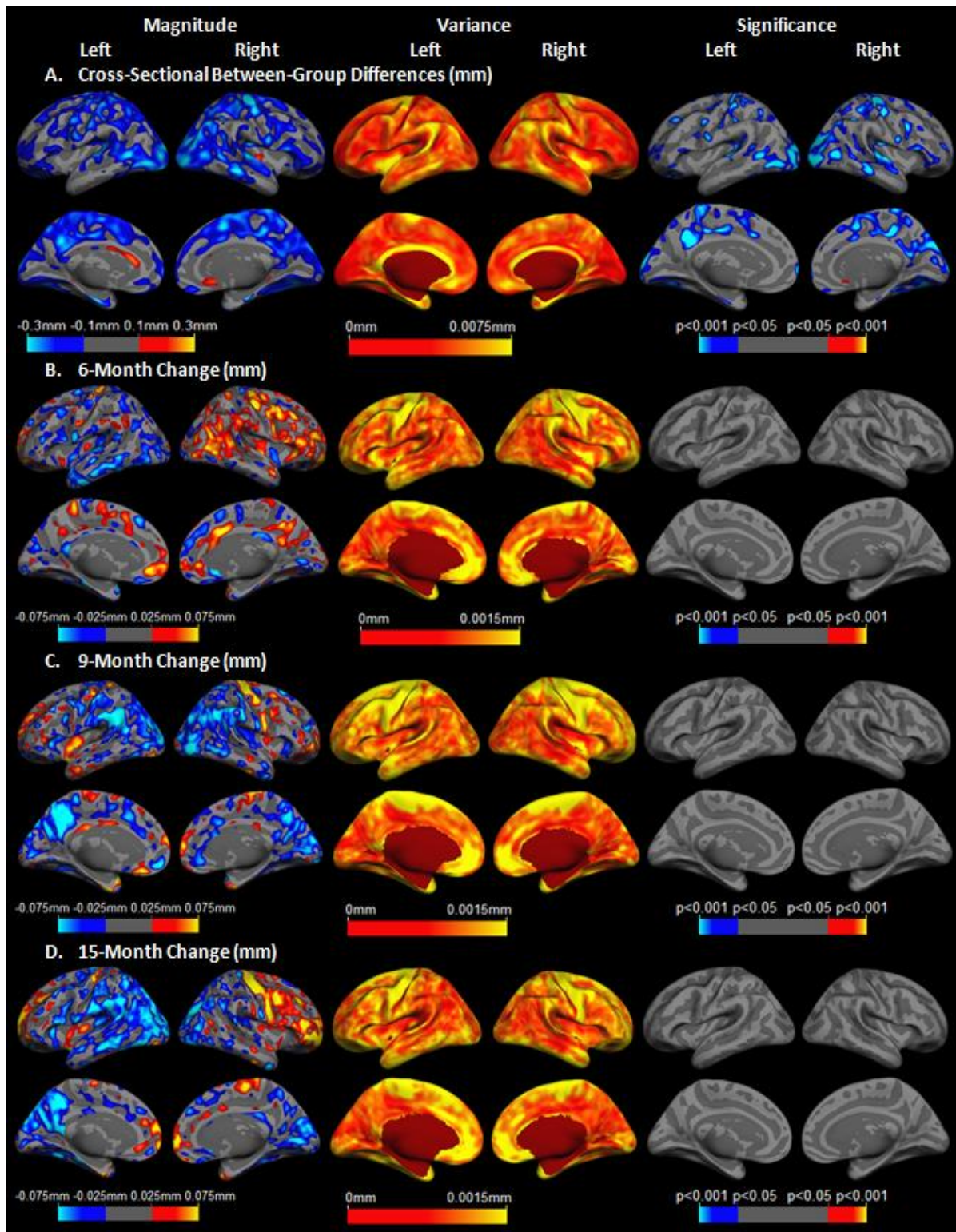


Figure 9-8. (Left) Magnitude, (middle) variance and (right) significance maps of between-group differences within the left and right hemispheres in: A) cross-sectional thickness (mm); B) rate of change over six months C) nine months and; D) 15 months. All analyses are adjusted for age, gender, study site and scan interval. Variance is calculated as the square of the standard error. The significance map colour bars represent the T values between $p < 0.001$ and $p < 0.05$ adjusted for multiple comparisons using FDR correction ($p < 0.05$).

Rates of 15-month change within all atlas regions are shown in Table 9-4, along with adjusted between-group differences and effect sizes (ES). Four of the six parietal regions showed significant ($p < 0.05$) or borderline significant ($p < 0.08$) between-group differences, as was the case for three of the five occipital regions, over 15 months. Of the frontal regions only the frontal pole showed a significant difference, with the control group undergoing more thinning in this region than the HD group. The middle temporal and parahippocampal gyri were the only temporal lobe regions nearing significance. The atlas regions within which HD-related cortical atrophy was most severe, and consequently exhibited the largest effect sizes, were the cuneus (ES = 0.786; 95% CI 0.264, 1.373), precuneus (0.511; 0.036, 1.007) and lateral occipital cortex ((LOC) 0.495; -0.126, 1.041). There was also a strong between-group difference in the frontal pole (-0.709; -1.299, -0.194) with the control group showing a significantly faster rate of atrophy in this region.

The same analysis over the 9-month interval detected significant between-group differences in the precuneus only (-0.030mm (95% CI -0.060, -0.001) $p = 0.046$; ES = 0.473 (95% CI 0.924, 0.037)). Borderline significant between-group differences were found in the cuneus (-0.024mm (-0.050, 0.001) $p = 0.064$; ES = 0.468 (1.016, 0.143)) and supramarginal gyri (-0.029mm (-0.060, 0.002) $p = 0.066$; ES = 0.397 (0.820, 0.114)). Over the 6-month interval the finding of reverse atrophy in the frontal pole of the HD group (0.005mm, SD 0.094) drove the only significant between-group difference (0.062mm (0.021, 0.104) $p = 0.003$; ES = -0.688 (-0.155, -1.199)).

Table 9-4. Rate of cortical thickness change (mm) over 15 months (adjusted for interval) extracted from all atlas regions, with adjusted between-group differences and effect sizes.

Atlas Region	Controls Mean (SD)	HD Mean (SD)	Adj. Diff. (95% CI)	P-Value	Effect Size (95% CI)
Parietal Lobe					
Parietal Lobe Average	-.012 (.066)	-.037 (.057)	-.024 (-.049 , .001)	0.062	.449 (-.092, .936)
Superior Parietal	-.010 (.084)	-.039 (.072)	-.027 (-.060 , .005)	0.102	.399 (-.174, .904)
Inferior Parietal	-.013 (.066)	-.039 (.060)	-.027 (-.052 , -.001)	0.039	.492 (-.080, 1.021)
Supra-marginal	-.020 (.062)	-.046 (.061)	-.025 (-.050 , .0003)	0.052	.438 (-.079, .934)
Postcentral	-.016 (.084)	-.032 (.058)	-.013 (-.043 , .018)	0.418	.216 (-.445, .729)
Precuneus	-.002 (.065)	-.032 (.061)	-.029 (-.056 , -.003)	0.029	.511 (.036, 1.007)
Occipital Lobe					
Occipital Lobe Average	.001 (.058)	-.028 (.048)	-.025 (-.048 , -.003)	0.024	.552 (.059, 1.076)
Lateral Occipital	-.008 (.078)	-.039 (.057)	-.028 (-.057 , .001)	0.060	.495 (-.126, 1.041)
Lingual	-.003 (.052)	-.019 (.061)	-.012 (-.035 , .011)	0.297	.214 (-.218, .625)
Cuneus	.007 (.065)	-.032 (.049)	-.037 (-.061 , -.013)	0.003	.786 (.264, 1.373)
Pericalcarine	.008 (.075)	-.022 (.067)	-.025 (-.055 , .005)	0.099	.388 (-.075, .937)
Frontal Lobe					
Frontal Lobe Average	-.035 (.078)	-.029 (.050)	.012 (-.016 , .039)	0.401	-.237 (-.899, .391)
Superior Frontal	-.038 (.091)	-.037 (.068)	.005 (-.028 , .039)	0.748	-.082 (-.658, .466)
Rostral Middle Frontal	-.042 (.081)	-.024 (.058)	.025 (-.004 , .054)	0.094	-.441 (-1.036, .135)
Caudal Middle Frontal	-.038 (.098)	-.044 (.068)	-.001 (-.036 , .035)	0.974	.009 (-.638, .600)
Pars Opercularis	-.026 (.069)	-.041 (.057)	-.011 (-.037 , .015)	0.419	.195 (-.322, .759)
Pars Triangularis	-.037 (.089)	-.032 (.061)	.009 (-.024 , .042)	0.574	-.159 (-.739, .503)
Pars Orbitalis	-.030 (.104)	-.023 (.077)	.017 (-.021 , .055)	0.377	.195 (-.322, .759)
Lateral Orbitofrontal	-.012 (.081)	-.029 (.064)	-.013 (-.044 , .019)	0.436	.194 (-.309, .744)

Medial Orbitofrontal	-.009 (.089)	-.020 (.067)	-.007 (-.040 , .027)	0.693	.102 (-.495, .638)
Precentral	-.051 (.131)	-.045 (.099)	.015 (-.032 , .063)	0.525	-.160 (-.826, .328)
Paracentral	-.041 (.122)	-.033 (.081)	.015 (-.029 , .058)	0.508	-.179 (-.835, .350)
Frontal Pole	-.058 (.125)	.003 (.108)	.074 (.026 , .122)	0.003	-.709 (-1.299, -.194)
Temporal Lobe					
Temporal Lobe Average	-.010 (.060)	-.027 (.066)	-.018 (-.044 , .009)	0.188	.280 (-.153, .769)
Superior Temporal	-.018 (.053)	-.035 (.058)	-.013 (-.035 , .009)	0.244	.237 (-.227, .662)
Middle Temporal	-.012 (.063)	-.037 (.056)	-.024 (-.049 , .001)	0.059	.435 (-.082, .920)
Inferior Temporal	-.007 (.079)	-.025 (.067)	-.019 (-.051 , .012)	0.222	.295 (-.188, .868)
Banks of the Superior Temporal Sulcus	-.021 (.062)	-.040 (.076)	-.017 (-.046 , .011)	0.233	.229 (-.199, .688)
Fusiform	-.002 (.063)	-.021 (.069)	-.020 (-.048 , .008)	0.170	.294 (-.171, .769)
Transverse Temporal	-.014 (.107)	-.042 (.101)	-.031 (-.073 , .012)	0.155	.315 (-.173, .760)
Entorhinal	-.004 (.134)	-.005 (.135)	-.003 (-.061 , .055)	0.911	.026 (-.498, .525)
Temporal Pole	-.017 (.119)	-.016 (.114)	-.003 (-.052 , .046)	0.901	.029 (-.479, .527)
Parahippocampal	.006 (.066)	-.019 (.087)	-.030 (-.062 , .002)	0.069	.351 (-.059, .782)
All values are adjusted for scan interval. Generalised least squares regression models were fitted to estimate the between-group differences in thickness change, adjusted for age, gender and study site. Effect sizes were calculated as the estimated absolute adjusted mean difference of the metric between HD participants and controls, divided by the estimated residual SD of the HD group. These are reported with bias-corrected and accelerated bootstrapped 95% CIs.					

Using the standard formula (Julious 2009), with 90% power and two-tailed $p < 0.05$, it is estimated that trials over 15 months of a 50% or 20% effective treatment in early-stage HD would require a sample size in each randomised group of: 276 (95% CI 73 to 24,148) and 1,724 (454 to 150,925) respectively for the occipital lobe, and 417 (96 to 9,931) and 2,606 (600 to 62,071) for the parietal lobe.

9.5.4 Discussion

Biomarkers of disease progression are required to facilitate efficacy testing of putative disease-modifying treatments in clinical trials. Using a commonly applied software package, FreeSurfer, and a large cohort of early manifest HD patients and controls, the sensitivity of cortical thickness analysis to longitudinal change was investigated over short and varying time intervals.

Consistent with previous studies of cortical thickness in HD, significant thinning of up to ~12.6% of the average control thickness was seen cross-sectionally over the majority of the cortical mantle (Nopoulos et al. 2010; Rosas et al. 2002; Rosas et al. 2008; Tabrizi et al. 2009). In contrast to other volumetric neuroimaging biomarkers in HD (Henley et al. 2006; Tabrizi et al. 2011) no significant between-group differences in atrophy were detectable over six, nine or 15 months. Several atlas regions in the parietal and occipital lobes, when extracted, did show significant or borderline significant atrophy in the HD group compared with controls. The finding of a significantly increased rate of atrophy in the control group compared with the HD group at the frontal pole is biologically implausible and most likely driven by measurement error. This could be due to artificially large thickness estimates at baseline in the control group caused by erroneous inclusion of dura and non-brain tissue; this is particularly common in control scans where the brain is large and therefore the cortical GM surface is near to the dura.

The highest 15-month effect sizes were found within the cuneus, precuneus and LOC; 0.786 (95% CI 0.264, 1.373), 0.511 (0.036, 1.007) and 0.495 (-0.126, 1.041) respectively. These effect sizes and 95% CIs are roughly comparable to unpublished data from this cohort (reported in Chapter 18), over 15 months, for two of the strongest cognitive tests: the SDMT (0.799 (95% CI 0.344, 1.254)) and the Letter Fluency task (0.664 (-0.031, 1.32)). This method is not however comparable to the strongest volumetric or diffusion neuroimaging biomarkers: for example the CBSI (1.191 (95% CI 0.742, 1.687)) and caudate AD (1.174 (0.839, 1.493)).

Cortical thinning as a biomarker for future clinical trials in HD is most pertinent for therapeutics targeting the cortex. Although the magnitude and statistical maps are of use for localising between-group differences during the natural course of disease, a parcellation approach per lobe is suggested here as the optimal use of cortical thickness data as a quantifiable biomarker. The focus of action of such a therapeutic would not be expected to localise to one atlas region therefore extracting averaged data from either the parietal or occipital lobe would be more appropriate. It is estimated that trials over 15 months of a 50% or 20% effective treatment in early-stage HD would require a sample size in each randomised group of: 276 (95% CI 73 to 24,148) and 1,724 (454 to 150,925) respectively for the occipital lobe, and 417 (96 to 9,931) and 2,606 (600 to 62,071) for the parietal lobe (Julious 2009). These sample size estimates and the CIs around them are very large, extending up to unfeasible sample requirements.

This study used the default FreeSurfer parameters and therefore validated optimisations and manual edits may enhance the reliability and sensitivity of the measures. Nevertheless these results should be repeatable and generalizable to other studies using the default parameters. The scope of this study did not include a direct comparison of the results between scanners or the assessment of clinical correlations with these findings. This should be investigated further with larger numbers and over varying disease stages.

In summary, this is the first study to assess cortical thinning over time in a relatively large early-stage HD cohort compared with controls and across multiple scanners; important for future large multi-site studies. Overall these findings suggest that parcellated regional thickness analysis (particularly average occipital or parietal lobe thickness), over longer time intervals may enrich information available from striatal and global biomarkers in HD, providing a more accurate representation of the natural progression of pathology across the whole-brain. These lobular averages will be included in a larger direct statistical comparison of biomarkers in HD (Chapter 18).

10 Development and Administration of a Novel Cognitive Test: Multi-Modal Emotion Recognition in HD

10.1 Background

Recognising emotions in others from their facial or vocal expressions and body language is a key social skill which is impaired in HD; the literature on which has been reviewed by Henley et al. (Henley et al. 2012). Impairment in this ability could lead to problems with social relationships and therefore a clearer understanding of the profile of this deficit may help guide clinical care. Additionally, a large observational study found recognition of negative emotions to be the only cognitive task in a comprehensive battery to be associated with disease progression in pre-manifest HD (Tabrizi et al. 2013) implicating this as a strong cognitive test to track the pathological development of disease. Investigation of emotion recognition deficits in HD may also help to clarify whether emotion recognition can be conceptualised as a single unified ability or whether there are subdivisions within this, either between emotions or stimulus modalities. For example, if someone can recognise one emotion very well can they recognise all emotions well? If someone is sensitive to emotion portrayed verbally are they also sensitive to visual displays of emotion?

On this basis a novel cognitive task aimed at furthering our knowledge of multi-modal emotion recognition in HD was developed and administered in an early-stage HD cohort; the details of which are described in this chapter.

There are six basic emotions, as outlined in much of the seminal work by Paul Ekman and colleagues (Ekman 1992): happiness, sadness, fear, surprise, anger and disgust. These six are thought to be cross-cultural, with a biological basis. Studies in healthy populations have found a low correlation between a person's ability to recognise positive and negative emotions, leading to the suggestion that these skills may be independent (Suzuki et al. 2014). Similar theories have proposed that emotion recognition may be a broad ability consisting of related, but partially dissociable skills involved in the recognition of positive and negative emotions (Hall 2001;Schlegel et al. 2012).

Studies in HD initially reported a disproportionate impairment of disgust recognition (Hayes et al. 2007;Sprengelmeyer et al. 1996;Wang et al. 2003). Later findings however suggested that other negative emotions (anger and fear) were equally, if not more affected (Calder et al. 2010;deGelder et al. 2008;Hayes et al. 2009;Henley et al. 2008;Ille et al. 2011b;Milders et al. 2003;Montagne et al. 2006;Snowden et al. 2008;Tabrizi et al. 2009). A more recent study with a much larger cohort detected significant impairment in recognition (from static photos) of both positive and negative emotions in HD (Labuschagne et al. 2013). Results from this higher powered study suggest that there is a smaller, but significant, effect on positive

emotion recognition also. These results are supportive of the concept of dissociable abilities in recognition of positive and negative emotions, with both affected but to varying degrees in HD.

It is unclear in the general population whether emotion recognition from different modalities can be explained by a unitary ability. There is some evidence that emotion recognition may be modality-specific as performance in a healthy cohort correlated poorly between tests in different modalities (Scherer and Scherer 2011). Conversely, Schlegel et al. (Schlegel et al. 2012) concluded that emotion recognition performance in a healthy cohort across stimulus modalities (audio, film, photo and audio-film) could be explained by a single ability dimension. There is therefore no firm agreement on this point.

In HD, impairments in disgust (Hayes et al. 2007), anger and fear (Calder et al. 2010; Snowden et al. 2008) recognition have been reported across both facial and vocal modalities. The pattern of the severity of emotion-specific deficits across these stimulus types however is unclear. Evidence from vocal stimuli suggests a different deficit to that reported from static facial stimuli, although this relationship was not directly tested (Robotham et al. 2011).

The profile of emotion recognition deficits across different stimulus modalities has important methodological implications. If a varying pattern of impairment is seen across stimulus modalities the use of one to reflect psychosocial functioning would not be appropriate or reflective of the deficit as a whole. If the deficit is statistically similar across stimulus modalities this would implicate a central unified emotion-specific, or negative emotion-specific, pathology underlying multi-modal recognition of that or those emotions.

Overall, the literature suggests that emotion recognition ability can be subdivided into that for positive and negative emotions. It seems that only negative emotion recognition is impaired in HD. It is however unclear whether emotion-specific recognition is consistent across stimulus modalities. The aim of this study was to directly compare, for the first time, the profile of HD-related emotion recognition deficits across different stimulus modalities (facial photos, vocal expressions and dynamic film clips of facial expressions). This is also the first study to investigate emotion recognition in HD from the arguably more ecologically valid dynamic facial film stimuli, with the aim to establish a more clinically-relevant profile of this impairment. Finally, in order to test whether the deficits are related to task demands, the HD group's performance will be compared to the performance of a matched control group. The pattern of errors made across emotion cues and stimulus modalities will be compared to the between-group differences to identify whether the most misidentified emotions were also the ones showing the largest between-group differences.

10.2 Task Design

10.2.1 Stimuli

This cognitive task was designed with three sections, each presenting different emotion stimuli:

- 1 Static black-and-white photos (the Manchester Face Set (Snowden et al. 2008;Whittaker et al. 2001)) – examples in Figure 10-1
- 2 Non-verbal vocal audio clips (Sauter et al. 2010a;Sauter et al. 2010b)
- 3 One-second film stimuli (Simon et al. 2008).

The Manchester Face Set (Snowden et al. 2008;Whittaker et al. 2001) is a modern variation on the widely-used Ekman face stimuli. The faces were all full-face frontal views of actors posing for a photograph. The vocal stimuli were taken from an emotion sounds test of recorded non-verbal sounds corresponding to happiness (laughter), sadness (sobbing), anger (growls), fear (screams), disgust (retching) and surprise (gasping) (Sauter et al. 2010a;Sauter et al. 2010b). The film stimuli were one-second colour film clips of drama students who were asked to produce each expression in about one second starting with a neutral face and ending at the peak of the expression (Simon et al. 2008). After the one second expression these film clips remained on the screen but static. In a previous study it was noted that 'disgust' in English can have both visceral and moral connotations (Snowden et al. 2008) and therefore variations in meaning might underlie performance differences for 'disgust' across studies involving different languages. Consequently, for consistency and comparability, all disgust stimuli in the current study were visceral in type.



Figure 10-1. Examples of static photo cues expressing happiness, anger, disgust and fear (top-bottom, left-right).

10.2.2 Paradigm

A forced-choice paradigm was used, instructing the participant to choose the emotion portrayed via a keyboard button press. Each cue was presented or played on a computer screen along with a list of possible emotion options: 1=anger, 2=disgust, 3=fear, 4=happiness, 5=sadness, 6=surprise. There were no time limits and only when a response was given did the task move to the next stimulus. Within each section (photos, vocal cues and film clips) the stimuli were presented one at a time in a pseudo-random order. Within each stimulus modality the six emotions were presented five times each, resulting in a total of 30 stimuli per stimulus modality and each emotion tested 15 times (across three modalities). The maximum score for the whole task therefore was 90. It should be noted that, for simplicity, each modality was run in blocks therefore there is a possibility that there may be a priming effect.

10.2.3 Control Tasks

Two additional control tests were performed: firstly, the Benton Facial Recognition Test (BFRT short-form (Benton 1980); a face-matching task) was administered in order to control for basic facial processing ability; secondly, a short motor control task was run. For this, one of the numbers 1-6 appeared on screen in a pseudo-random order and, using the same keyboard buttons as the main task, the participant was asked to press the corresponding button as quickly as they could. Twelve numbers were presented in total. The average time (ms) taken for the button response, excluding the first number, was calculated and used as a control (referred to as motor response time) for motor impairments in statistical analyses.

10.3 Application to an HD Cohort

Participants were recruited at the London site of the PADDINGTON study (Hobbs et al. 2013). HD participants (n=15) were in stage I of disease based on their TFC score from the UHDRS (1996), with an average disease burden score of 368.61 (Penney et al. 1997). Fifteen of the 18 control participants were gene-negative siblings or spouses of the HD participants. This was deemed to be preferable to using unrelated healthy controls in order to attempt to match for social and environmental factors. Participant characteristics are summarized in Table 10-1. One control participant did not complete the BFRT due to time constraints.

Table 10-1. Emotion recognition task participant demographics.				
	Controls		HD	
	n	Mean (SD) min - max	n	Mean (SD) min - max
Age (years)	18	56.31 (8.92) 43.74-77.71	15	52.29 (9.41) 41.16-68.03
Gender (M/F)	11/7		3/12	
CAG repeat length		N/A	15	42.87 (2.56) 39.00-46.00
TMS		N/A	15	19.60 (10.10) 7.00-45.00
Disease Burden ^a		N/A	15	368.61 (103.94) 232.60-562.98
Motor response time (ms) ^b	18	1021.74 (182.70) 792.18-1487.88	15	1368.22 (313.23) 952.41-2021.18
Benton Facial Recognition Test (BFRT) score ^c	17	48.94 (3.33) 43.00-57.00	15	43.00 (4.61) 36.00-51.00
Education level (ISCED)	18	3.88 (1.40) 1.00-6.00	15	3.27 (1.22) 2.00-5.00
^a Disease burden formula (Penney et al. 1997): (CAG-35.5) × age; ^b motor response time was assessed and used as a covariate in later analyses to remove a potentially confounding motor component from the main task, the methods for which are described in Section 10.2.3; ^c BFRT (Benton 1980) score adjusted for age and years of education. ISCED = International Standard Classification of Education.				

10.4 Statistical Analysis

All statistical analyses were conducted in STATA v12. The BFRT was analysed using generalised least squares regression, allowing for different variances in controls and HD participants and adjusting for age, gender, education level and motor response time. Linear regression, with modifications to allow for the non-continuous and non-independent nature of the outcome variable, was used to analyse the data from the emotion recognition task. The outcome variable in the model was the number of errors (out of five) for each stimulus modality-emotion combination (e.g. photos of fear, film clips of anger etc.). Predictors were a three-way interaction between disease group, stimulus modality and emotion, with adjustment for age, gender, BFRT score, education level and motor response time. To simultaneously allow for the non-normality of the outcome, and for the correlation between the 18 scores for each subject, non-parametric bias-corrected 95% and 99% bootstrap CIs (Efron and Tibshirani 1993) were computed from 2000 bootstrap samples, clustered by subject and stratified by group. As such, p-value accuracy was reported to > or <0.05 or <0.01. For the same reason, the three-way interaction was tested with a permutation test. Disease status (control or HD) was permuted 10,000 times, with the Wald test statistic for the interaction term recorded each time. The proportion of times the test statistic was more extreme than the observed statistic from the model was then computed to give the p-value.

10.5 Results

10.5.1 Facial Recognition

Despite the HD group mean of 43 on the BFRT being within the performance range classified as ‘normal’ (41-54), this was, on average, significantly worse than controls (estimate -6.76 (95% CI -9.80, -3.71);

$p < 0.001$). A negative association between BFRT score and emotion recognition errors was also observed. This supported the decision to adjust for BFRT scores in the main analyses in order to exclude the possibility that any between-group differences in emotion recognition merely reflected differences in facial recognition.

10.5.2 Emotion Recognition

Table 10-2. Mean (SD) number of errors in both the control and HD groups for: each emotion within each stimulus modality (/5; e.g. photos of fear); each emotion across all stimulus modalities (/15; e.g. fear recognition from photos, vocal and film cues); total within each stimulus modality across all emotions (/30; e.g. total number of errors from all photo stimuli); and the total overall (/90).

Mean Number of Errors (SD)								
Group	Stimulus Modality	Fear	Surprise	Sadness	Anger	Happiness	Disgust	Total
Control	All	4.56 (2.96)	3.94 (1.86)	2.94 (1.83)	2.44 (1.54)	1.56 (1.25)	0.72 (0.83)	16.17 (5.75)
	Photo	1.72 (1.18)	1.94 (1.61)	0.61 (0.61)	0.94 (0.94)	0.44 (0.62)	0.33 (0.49)	6.00 (3.03)
	Vocal	1.33 (1.33)	1.28 (1.23)	1.67 (0.97)	1.39 (0.98)	1.06 (1.21)	0.22 (0.43)	6.94 (3.10)
	Film	1.50 (1.15)	0.72 (0.75)	0.67 (0.97)	0.11 (0.32)	0.06 (0.24)	0.17 (0.38)	3.22 (1.52)
HD	All	8.67 (2.94)	5.87 (1.73)	5.07 (2.76)	7.87 (3.89)	2.87 (1.68)	4.93 (3.39)	39.87 (20.04)
	Photo	3.20 (1.21)	2.47 (1.25)	1.60 (1.40)	2.80 (1.57)	0.73 (0.88)	1.73 (1.49)	12.53 (4.56)
	Vocal	3.07 (1.33)	2.60 (1.40)	2.40 (1.12)	2.40 (1.30)	1.93 (1.58)	1.53 (1.25)	13.93 (4.83)
	Film	2.40 (1.24)	0.80 (0.77)	1.53 (1.41)	2.67 (1.84)	0.20 (0.56)	1.67 (1.40)	9.27 (4.27)

The observed means and SD in the number of errors are presented in Table 10-2, stratified by group, emotion and modality. Of all six emotions tested, fear was the most misidentified whilst happiness and disgust were recognised most easily in both the control and HD groups. Results of the permutation test showed that the three-way interaction between group, modality and emotion was statistically significant at the 5% level ($p = 0.013$) and as such the adjusted between-group differences in expected number of emotion recognition errors are presented separately for each emotion modality combination in Figure 10-2.

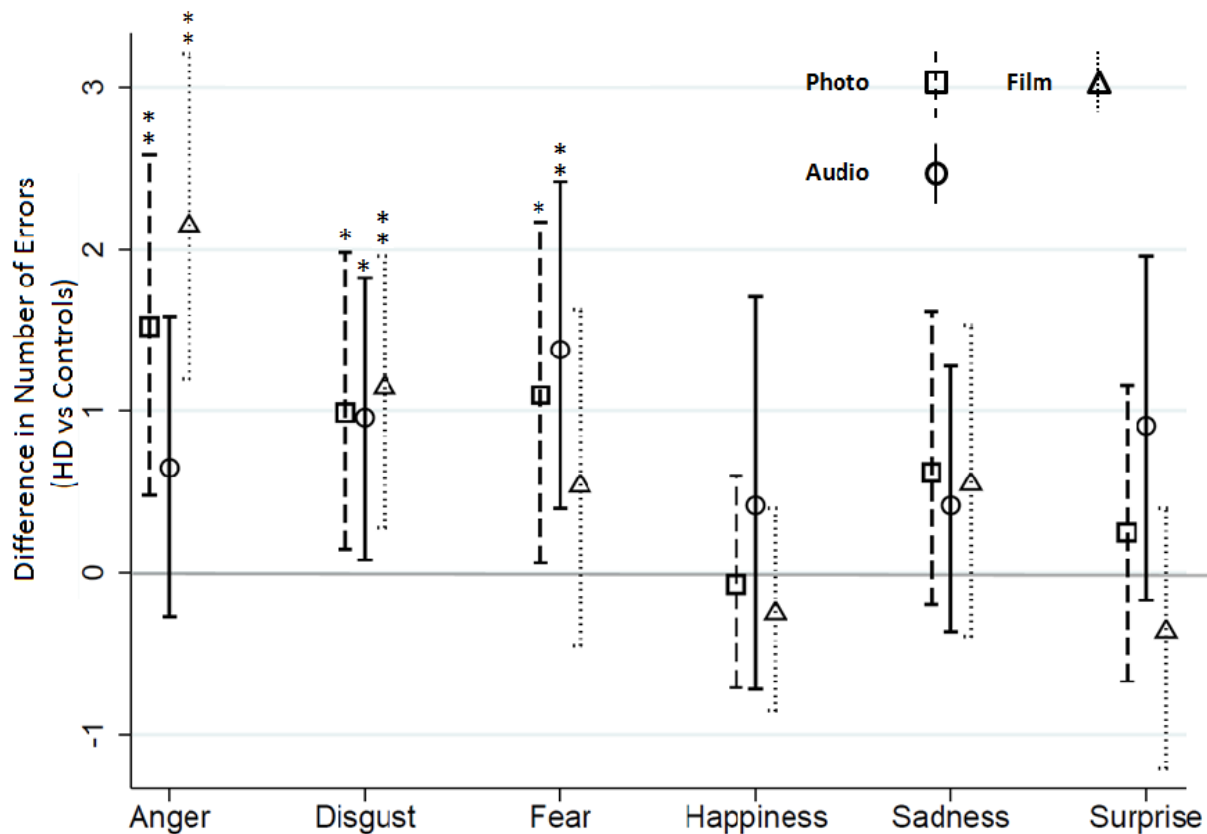


Figure 10-2. Estimated between-group differences (HD vs control) in number of errors out of five for each emotion modality combination, with 95% bootstrapped bias-corrected and accelerated confidence intervals. Positive values indicate that HD participants made more errors than controls. Statistical significance is highlighted with: * $p<0.05$; ** $p<0.01$.

For photos the largest difference between HD and control group performances was seen for anger, with those for fear and disgust also achieving statistical significance. For vocal stimuli the largest difference between HD and control group performances was for fear with the difference for disgust also statistically significant. The difference for anger favoured controls but was smaller than that for photos and film although the CI was relatively wide. For film, as with photos, the largest difference was seen for anger, with that for disgust also achieving statistical significance. Here the between-group difference for fear was smaller than seen with the other stimulus modalities although again the difference still favoured controls with the CI being reasonably wide. For happiness, sadness and surprise none of the differences between the HD and control groups achieved statistical significance although in some cases differences were relatively large, albeit with wide CIs.

Of the errors in the HD group the most common mistakes (reported as a percentage of the total number of responses for the emotion displayed) were: anger mistaken for disgust (25%, i.e. 25% of the responses made by the HD group to anger stimuli involved the incorrect labelling of these as disgust); disgust

mistaken for sadness (10.3%); fear mistaken for surprise (30.9%); happiness mistaken for surprise (11.1%); sadness mistaken for disgust (12.4%); surprise mistaken for happiness (16.1%). The percentages of each response made within each stimulus modality are illustrated in Table 10-3. From this we can see that the results are suggestive of two groupings within which the majority of misidentifications occur in this HD cohort: 1) anger, disgust and sadness; 2) fear, happiness and surprise.

Table 10-3. A matrix of the percentage of answers given by the HD group for each emotion displayed, separated by stimulus modality.																			
		Emotion Displayed																	
		Anger			Disgust			Fear			Happiness			Sadness			Surprise		
		Photo	Vocal	Film	Photo	Vocal	Film	Photo	Vocal	Film	Photo	Vocal	Film	Photo	Vocal	Film	Photo	Vocal	Film
Answer Given	Anger	44.0%	51.4%	46.7%	9.3%	6.8%	10.7%	5.3%	9.5%	8.0%	1.3%	1.3%	0.0%	5.3%	8.0%	4.0%	2.7%	2.7%	4.0%
	Disgust	18.7%	13.5%	42.7%	65.3%	68.9%	66.7%	9.3%	14.9%	10.7%	0.0%	1.3%	0.0%	10.7%	10.7%	16.0%	8.0%	13.5%	2.7%
	Fear	5.3%	16.2%	4.0%	4.0%	5.4%	5.3%	36.0%	39.2%	52.0%	1.3%	2.7%	1.3%	4.0%	5.3%	8.0%	5.3%	9.5%	9.3%
	Happiness	1.3%	2.7%	0.0%	5.3%	4.1%	1.3%	0.0%	1.4%	0.0%	85.3%	61.3%	96.0%	4.0%	18.7%	1.3%	25.3%	23.0%	0.0%
	Sadness	21.3%	10.8%	4.0%	10.7%	6.8%	13.3%	5.3%	14.9%	1.3%	0.0%	12.0%	2.7%	68.0%	52.0%	69.3%	8.0%	2.7%	0.0%
	Surprise	9.3%	5.4%	2.7%	5.3%	8.1%	2.7%	44.0%	20.3%	28.0%	12.0%	21.3%	0.0%	8.0%	5.3%	1.3%	50.7%	48.6%	84.0%

10.6 Discussion

This was the first study to compare the profile of the emotion recognition deficit in HD across different stimulus modalities, including the first use of dynamic film stimuli of facial expressions. It is thought that recognition of emotions of positive and negative valence constitutes dissociable abilities but it is currently unclear whether emotion recognition across different stimulus modalities should be conceptualised as a unitary ability. The consistent finding of impaired negative emotion recognition in HD is supportive of the theory of positive and negative subdimensions in emotion recognition ability. This deficit is independent of individual stimulus or task demands as the pattern of errors made did not match the pattern of HD-related deficit in comparison to the control group performance. This valence-specific emotion recognition ability was found in this study to be consistent across stimulus modalities but the emotion-specific profile of the impairment differed significantly. This could be explained by the theory of related yet specific skills that incrementally contribute to emotion recognition ability (Hall 2001;Schlegel et al. 2012). These findings show that the use of emotion recognition performance from one stimulus modality is not representative of the deficit as a whole and therefore is not fully reflective of psychosocial functioning. These findings could have implications in terms of designing future tests and care giving.

There does not seem to be a clear link between the number of errors and the pattern of impairment in HD participants compared with controls. For example, disgust was the most accurately identified emotion by the control group and the second best only behind happiness in the HD group, yet recognition of this emotion was consistently different between groups in all stimulus modalities. These findings do not support the theory that the disproportionate emotion-specific deficits found in HD are a result of increased cognitive demand, test cue difficulty or expression complexity but instead are indicative of emotion-specific or valence-specific cognitive pathology.

The significant three-way interaction between disease group, stimulus modality and emotion suggests however that impairment may not be entirely consistent across stimulus modality and emotion. The largest between-modality difference was shown between film and vocal anger recognition; anger being recognised by the HD group at a closer level to controls from the vocal cues than the film clips. In addition impairment in fear recognition was estimated to be greater than that for anger in the vocal stimuli, whereas the converse was true for film stimuli. If these observed between-modality differences in emotion recognition deficits in HD were to be replicated in a larger sample, this finding could have implications in terms of designing future tests and care giving. For example, emotion recognition performance from the traditional static photo modality does not appear to be as clinically relevant as previously assumed; in order to establish a more accurate profile of this impairment in everyday life a multi-modal test of emotion recognition is recommended. In terms of care giving these results suggest that educating care-givers of this

potential deficit and training in the use of a varied communication style, incorporating both visual and vocal expressions, may facilitate emotion recognition and consequently social functioning.

From the mistakes being made we can see that there are two groupings of emotions that are being confused in HD: 1) anger, disgust and sadness; 2) fear, happiness and surprise. These groupings are consistent with the schematic representations of the relations between different facial expressions proposed by Woodworth & Schlosberg (Woodworth and Schlosberg 1954) and Calder et al.'s (Calder et al. 1996) hexagonal continuum of emotions: happiness to surprise to fear to sadness to disgust to anger to happiness. The two groupings can be thought of as open and closed expressions: with anger, disgust and sadness characterised by a closing of the eyes and lowering of the eyebrows; and fear, happiness and surprise characterised by an opening of the eyes and raising of the eyebrows. These similarities between expressions increase the demands on perceptual processing when discriminating between emotions within a 'set'. In terms of complexity, disgust and anger also have a conceptual overlap and this may be a factor behind anger-disgust mistakes.

In the HD group the error rates in the photo, vocal and film stimuli were 41.8%, 46.4% and 30.9% respectively. Emotion recognition from film cues (the least effected modality) may be more reflective of typical social interactions than static photos or vocal cues and therefore performance here is arguably of more clinical relevance. Although disgust, anger and fear recognition combined across stimulus modalities was significantly impaired in the HD group compared with control group performance these responses were erroneous just 32.9%, 52.5% and 57.8% of the time respectively; levels indicative of substantial remaining functional capacity in this early-stage HD cohort.

Previous studies of gene-negative siblings or spouses of HD participants have found impairments in facial anger recognition compared with unrelated healthy controls (Gray et al. 1997; Sprengelmeyer et al. 2006). Our control group (15 of the 18 were gene-negative siblings or spouses of HD participants) found anger to be only the fourth (out of six) most difficult emotion to recognise when errors from all stimulus modalities were combined (2.44 mean errors (SD 1.54) out of 15 responses). This does not support the idea of anger recognition impairments in gene-negative people from an HD family although this was not done in comparison to an unrelated healthy control group.

This study was limited by sample size and the exploratory findings would clearly benefit from replication. Additionally, medication was not taken into account in this study and recent evidence suggests that neuroleptics and serotonin reuptake inhibitors may have a significant effect on emotion recognition (Labuschagne et al. 2013). However, even if medication effects are significant in this cohort this is still a clinically relevant impairment and there is no evidence to suggest that this would differentially affect specific emotions or modalities. Consequently the patterns of the deficit reported here would not be

hypothesized to change with adjustment for medication usage. Finally, due to the clinical population available to us there was an imbalance in males and females within the two groups. This was accounted for in all statistical analyses by adjusting for gender. The fact that the behavioural results corroborate previous work suggests that this sample is likely to be representative and that these novel, albeit tentative, findings are worthy of follow-up.

In conclusion, the direct comparison of emotion recognition deficits across multiple stimulus modalities make this study a thorough investigation of several aspects of emotion recognition in HD. Consistent with previous reports, anger, disgust and fear recognition was shown to be impaired in HD. There was however evidence of differences in the extent of impairment relative to controls for specific emotions between the traditional static photo modality, vocal stimuli, and the more clinically relevant film clips, never before used in HD. Impairment does not seem to be due to task demands or expression complexity as the pattern of between-group differences did not correspond to the pattern of errors made by either group, therefore implicating emotion-specific cognitive processing pathology. These findings may have implications for future test design and care giving.

11 PART 1: A Summary

In Part 1 of this thesis regional volumetric image analysis protocols and a novel cognitive test were developed, tested and, for the volumetric biomarkers, compared to automated alternatives. Another automated methodology, in this case FreeSurfer cortical thickness analysis, was assessed for its suitability as a biomarker in HD and applied to the PADDINGTON data.

The caudate and putamen of the striatum are central to HD pathology (Vonsattel & DiFiglia 1998) and are therefore strong candidates as biomarker targets in HD. Although shown in post-mortem data to be similarly affected by disease (Aylward et al. 2004; Georgiou-Karistianis et al. 2013b; Halliday et al. 1998), the volumetric caudate biomarkers were reported (in Chapter 7) to be substantially more sensitive to atrophy than the putamen biomarkers. This is most likely due to the poorer definition of the putamen boundary. Automation of volumetric biomarkers is required in order to facilitate large-scale volumetric analysis in upcoming clinical trials. Several automated methods emerged from the method comparison as viable alternatives to the current manual and semi-automated gold-standards; most notably STEPS and FreeSurfer. Based on these results it is suggested that validated automated segmentation could replace manual caudate delineation at baseline for computation of the CBSI, fully automating this gold-standard technique.

The cerebellum is an under-investigated region in HD and therefore methods for analysing cerebellar volume were developed and tested in Chapter 8. These methods, most likely due to the complex border morphology of the cerebellum, were found to be substantially affected by inconsistencies in scan acquisitions and poor at detecting longitudinal change. The manual delineation protocol, developed in Section 8.3, was found to be sufficiently reliable on cross-sectional data and deemed to be the closest of these methods to a robust gold-standard; based on the inherent advantage of human judgement over automated methods.

The use of FreeSurfer software to assess cortical thinning over time in HD has recently been published, for the first time, as a biomarker in a clinical trial (Rosas et al. 2014). There is however currently very little in the HD literature testing the suitability of this technique as a biomarker. Exploratory investigations into several FreeSurfer parameters were conducted as pilot tests of this software in HD and reported in Section 9.4. The results reveal a process that can be biased by parameter selection. It is therefore imperative that there is clear *a priori* stipulation of the parameters, software version and statistical analysis to be used in order to facilitate unbiased analysis, replication and verification of any results published.

With *a priori* defined parameters FreeSurfer analysis was conducted on the PADDINGTON data and reported in Section 9.5. No significant between-group differences in atrophy rates were detected over 6-, 9- or 15-month intervals from statistical maps. However, when average thicknesses within atlas regions

were extracted, significant atrophy was detectable within several predominantly occipital and parietal regions. A parcellation approach per lobe is therefore suggested as the optimal use of cortical thickness data as a biomarker. These metrics were therefore included in a comprehensive biomarker battery to directly compare this technique against other, more widely used, neuroimaging biomarkers; the results of which are reported in Chapter 18.

The development and application of a novel test of multi-modal emotion recognition in HD was reported in Chapter 10. The finding of a differential pattern of impairment in emotion recognition between different stimulus modalities (photo, vocal and film stimuli) may have implications for future test design and care giving; particularly based on the findings from the more ecologically valid film stimuli, used here for the first time in HD.

Part 2 of this thesis reports results from exploratory investigations of neuroimaging associations with clinical and cognitive symptoms in HD. This builds upon the technical developments in Part 1 by applying several of the methods to large HD cohorts.

12 PART 2. Exploratory Investigations of Clinical, Cognitive and Neuroimaging Associations in HD

Despite the clear relevance of brain changes to the neuropathology of HD there is a need for a more comprehensive understanding of the clinical and cognitive implications of specific neuroimaging findings. As indirect measures of symptom progression, an enhanced knowledge of the relationship between neuroimaging metrics, brain functioning and behaviour will also serve to strengthen the argument for the clinical relevance of neuroimaging as surrogate outcome measures in future clinical trials. Consequently, data from the large TRACK-HD and PADDINGTON study cohorts was used in several exploratory investigations of clinical, cognitive and neuroimaging associations in HD; introduced below and the results from which are reported in Part 2 of this thesis.

Chapter 8 described the development and evaluation of a protocol for volumetric analysis of the cerebellum. Using this protocol, Chapter 13 reports the results of an investigation into the role of the cerebellum in HD. Volumetric and diffusion characteristics were compared between participants with early-stage HD and healthy controls. Subsequently, associations between disease-related cerebellar changes and clinical metrics were assessed in order to identify whether this under-investigated region has a larger impact on clinical presentation in HD than currently credited.

In order to more fully understand the emotion recognition deficit characterised by the novel cognitive test developed and applied in Chapter 10, an exploratory imaging investigation was conducted and is reported in Chapter 14. This was the first study to test associations between emotion recognition performance and micro-structural imaging metrics in HD.

Finally, results from FreeSurfer software's cortical thickness analysis (assessed in Chapter 9) are reported in Chapter 15 in an investigation of whether, and to what extent, thickness of the visual cortex affects cognitive task performance in the absence of any apparent visual deficits; using HD as the model. In addition to furthering our understanding of occipital lobe function, it was hoped that this study's findings would also enhance our understanding of cognitive deficits in HD.

13 Cerebellar Abnormalities in HD: A Role in Motor and Psychiatric Impairment?

13.1 Introduction

The cerebellum plays an important role in motor control (D'Angelo 2011). Motor signs of cerebellar dysfunction include: a wide-based ataxic gait, incoordination, oculomotor dysfunction (especially nystagmus), dysarthria and hypotonia. Emerging literature also implicates a role for the cerebellum, interacting with the basal ganglia, in dystonia (Neychev et al. 2008; Prudente et al. 2014); abnormal repetitive or twisting movements and postures, resulting from involuntary muscle contractions. In addition to these motor signs the cerebellum has been implicated in learning, cognition, attention, emotion (Allen and Courchesne 1998; D'Angelo 2010; Muller et al. 1998; Scharmuller et al. 2013; Schmahmann and Caplan 2006) and several psychiatric disorders (Kutty and Prendes 1981; Lauterbach 1996; Starkstein et al. 1988). Cerebellar disorders have informed our understanding of function, for example: a rare Cerebellar Cognitive Affective Syndrome (Schmahmann and Sherman 1998), caused by isolated cerebellar lesions, manifests as executive, visuospatial and linguistic dysfunction, accompanied by personality change; and inherited cerebellar ataxias manifest with cognitive impairment even in forms confined to the cerebellum, such as SCA13 or 14 (Durr 2010). Additionally, several toxins including alcohol and gabaergic medications are known to lesion the cerebellum and manifest as ataxia (Manto 2012). Alcoholic Cerebellar Degeneration has also been linked to cognitive and emotion deficits (Fitzpatrick and Crowe 2013).

Cerebellar symptoms and signs have been reported in HD. The most severe neuropathology in HD is focused within the basal ganglia and direct anatomical connections are known to exist between the basal ganglia and the cerebellum (for a review see (Wu & Hallett 2013)), along which abnormal activity could propagate with negative consequences (Bostan & Strick 2010). Given this evidence it is surprising that in adult-onset HD the cerebellum has not been investigated in more detail.

Although notably spared in comparison to other brain regions (Rosas et al. 2003), there is evidence that cerebellar abnormalities are present in HD. Autopsy reports of the cerebellum at end-stage disease suggest that the GM of the cortex and the deep cerebellar nuclei (most notably within the fastigial nucleus) are the most severely atrophied tissues (Jeste et al. 1984; Rodda 1981; Rub et al. 2013). *In vivo*, structural MRI studies have reported reduced cerebellar volume in manifest HD (Fennema-Notestine et al. 2004; Ruocco et al. 2006; Scharmuller et al. 2013). One study reported the cross-sectional percentage volume difference between the HD and control groups to be -3.7% and -16.9% in the GM and WM respectively (Fennema-Notestine et al. 2004). It should be noted however that this WM value includes the small deep GM nuclei as these cannot be distinguished from WM on T1-weighted MR images. Longitudinal atrophy of the cerebellum has been detected over one year (Ruocco et al. 2008) and two years (Hobbs et al. 2010b) in

early manifest disease; again the cerebellar cortex was found to be relatively preserved, whilst there were extensive losses in the WM and the brainstem. However, another study in stage I and II HD did not detect significant cerebellar atrophy in the HD group compared with controls but did find significant associations between cerebellar volume, disease duration and TFC (Rosas et al. 2003), a component of the UHDRS (1996). It is unclear at what stage of the disease process this atrophy emerges. Only one study in preHD has detected significant cerebellar volume loss (Gomez-Anson et al. 2009) whilst others have not (Hobbs et al. 2010b; Tabrizi et al. 2009).

In summary, inconsistent reports of the cerebellum in HD have emerged from the literature; autopsy data suggests atrophy is most marked within the cortical GM and GM of the deep cerebellar nuclei whilst structural imaging data (VBM) localises the most significant volume changes to the cerebellar WM or detects no disease effects. Furthermore, knowledge of the associations between cerebellar abnormalities and clinical signs in HD is limited. Consequently, a more focused, multi-modal imaging analysis of the cerebellum in HD was required; to-date no diffusion analysis focusing on the cerebellum has been published in HD. Furthermore, this study aimed to explore a possible role of cerebellar pathology in HD symptomatology.

13.2 Methods

13.2.1 Cohort

Data from 22 HD patients and 12 controls were taken from two sites (London and Paris) of the larger PADDINGTON study (Hobbs et al. 2013) provided the data were of good quality and there was complete cerebellum coverage in the diffusion image. HD patients were in stages I (n=20) and II (n=2) of disease based on their TFC score, indicating good capacity in functional realms. The majority of controls were spouses of the HD participants (n=7), with the aim to match for social and environmental factors. The remaining controls were gene-negative siblings (n=2) or unrelated volunteers (n=3). All subjects gave written informed consent and ethical approval was given by the Central London REC 4 and the CPP Ile de France VI.

13.2.2 Image Acquisition

3T MRI data (T1- and diffusion-weighted) were acquired at two study sites using Siemens Verio (Paris) and Siemens Tim Trio (London) scanners. T1 acquisitions were identical and diffusion acquisition parameters were carefully calibrated and tested to ensure that data collection was as consistent as possible (Muller et al. 2013) - details of all parameters are provided in Section 3.4. Data were pseudo-anonymised and archived on a secure web-portal. QC was performed on all scans; checking for coverage of the cerebellum, artefacts (e.g. movement, intensity) and sufficient tissue contrast for analysis. Scans passing these criteria were selected for analysis. All scans with sufficient coverage of the cerebellum passed visual QC.

13.2.3 Image Pre-Processing

T1-weighted scans were bias-corrected using the non-parametric non-uniform intensity normalization (N3) method of Sled et al. (Sled et al. 1998), with optimised parameters for 3T data as outlined in Boyes et al. (Boyes et al. 2008).

FA, MD, RD and AD maps were extracted from the tensors using the Camino software package (<http://www.cs.ucl.ac.uk/research/medic/camino>) and masked with a brain extracted (BET (Smith 2002)) average B0 image to remove non-brain tissue. Again using Camino, diffusion-weighted images were pre-processed with an initial affine registration to an averaged B0 reference image to correct for motion and eddy current distortions, and the gradient vectors were updated accordingly. Non-linear least squares regression was used to fit the tensors to avoid erroneous negative eigenvalues.

13.2.4 Image Analysis

Semi-automated delineation of the cerebellum was conducted in MIDAS software (Freeborough et al. 1997); example in Figure 13-1A. Details of this protocol development constitute Chapter 8 and the full protocol is included in Appendix Section 23.5.2. Briefly, scans were registered to MNI305 standard-space. Voxel intensity thresholds were set at 70% and 160% of mean brain intensity and a ‘seed’ was manually placed within the structure. These thresholds created an automatic initial outline which was then manually edited using landmark-defined cut-offs where appropriate: the cerebellum was manually separated from the cerebrum and borders were edited firstly in the coronal then sagittal views; the cerebellum-brainstem cut-off was defined, in the sagittal view, by a straight cut from the most anterior superior cerebellar GM to the most anterior inferior GM. Repeated analysis on a subset of scans with a week-long interval was found to output volumes with differences of <0.27%.

TIV was estimated according to a previously reported protocol (Whitwell et al. 2001) and cerebellar volumes are reported as a proportion of TIV to correct for inter-individual variation in head-size.

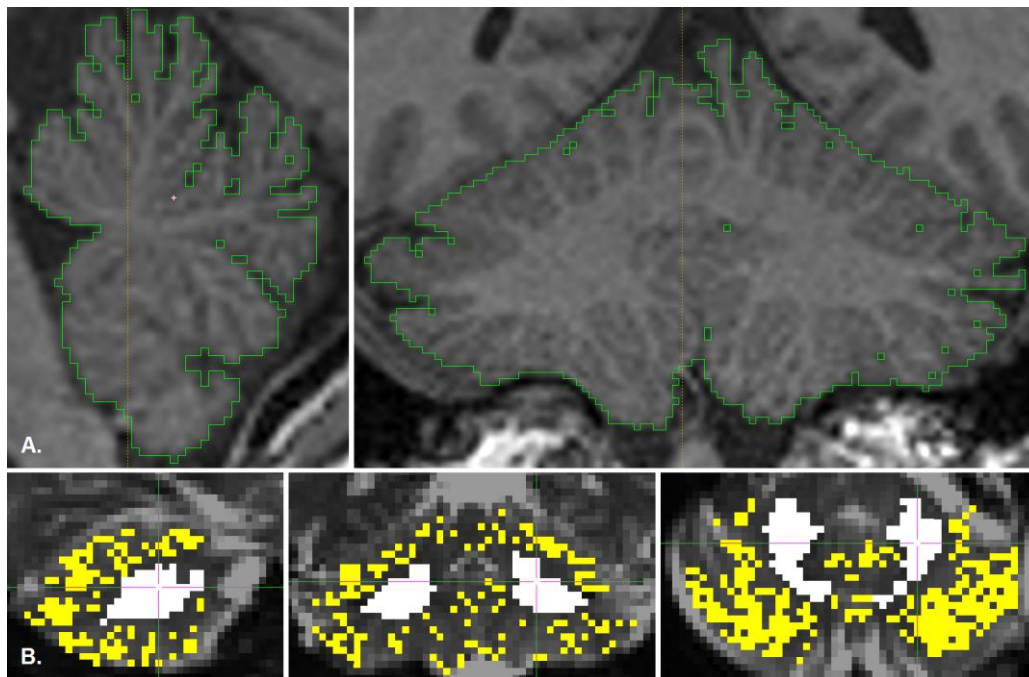


Figure 13-1. Examples of: A) a volumetric cerebellum delineation in MIDAS software; B) GM (yellow) and WM (white) masks for the diffusion analysis overlaid on the corresponding average B0 image.

Two further structural segmentations (tight and loose WM delineations) were conducted in MIDAS software for use as WM and GM masks for the diffusion analysis. Tight WM segmentations were performed by increasing the lower threshold of the original segmentations to 112% of the mean brain intensity and manually editing out GM regions included by these thresholds. The lower intensity threshold was set at 98% for the loose WM delineations. The whole-cerebellum, tight and loose WM segmentations were then converted to binary masks and the whole-cerebellum and tight WM masks were eroded by one voxel, reducing contamination from PVEs. The loose WM mask was subtracted from the eroded whole-cerebellum mask to provide a GM mask.

Structural scans were registered to the average B0 images using an initial affine step, followed by a non-linear registration approach (both using Nifty Reg software; <http://sourceforge.net/projects/niftyreg/>). All structural masks were then re-sampled, using the calculated parameters, onto the B0 maps using nearest neighbour interpolation (Figure 13-1B). Average FA, MD, AD and RD metrics were extracted from within these masked regions.

13.2.5 Voxel-Based Morphometry (VBM)

A voxel-wise structural analysis was conducted, in SPM8 on a Matlab R2012b platform, to examine between-group volumetric differences. Full details of the VBM analysis procedure were covered in Section 4.2.9. Briefly, using Unified segmentation (Ashburner & Friston 2005) and a study-specific DARTEL template (Ashburner 2007) between-group differences in GM and WM volume were assessed at the voxel-level followed by adjustment for multiple comparisons using FDR correction at the $p < 0.05$ level.

13.2.6 Motor and Psychiatric Assessments

Choice of clinical scales for the association analysis was based on previous literature linking motor (D'Angelo 2011) and psychiatric (Kutty & Prendes 1981;Lauterbach 1996;Starkstein et al. 1988) signs to cerebellar functioning. The UHDRS (1996) motor components tested were: TMS, saccade initiation, pronate/supinate-hand task, tandem walk, gait, retropulsion-pull test and finger tapping. Scores from left and right eyes/hands/fingers were added for saccade initiation, pronate-supinate-hand task and finger tapping. Anxiety, depression and irritability were quantified using a composite psychiatric morbidity score (HADS-SIS) from the Hospital Anxiety and Depression Scale (HADS (Zigmond & Snaith 1983)) and the Snaith Irritability Self-assessment scale (SIS (Snaith et al. 1978)).

13.2.7 Statistical Analysis

To investigate between-group differences in the imaging metrics generalised least squares regression was used, allowing for the variances to differ by group. This was due to an *a priori* belief that between-subject variation would be larger in the HD group than controls. Ordinary least squares regression models were fitted to investigate associations between clinical variables, volume and diffusion metrics within the HD group. All models (including the VBM) were adjusted for age, gender, study site and alcohol intake history (no/previous/current abuse). The VBM was additionally adjusted for TIV. Although some clinical outcomes take discrete values they were treated as continuous due to their ordinal nature and lack of data in some levels. If models showed signs of deviation from the normality assumption, bias-corrected bootstrap CIs were calculated based on 2000 replications, and p-values reported as < or > 0.05.

A post hoc investigation of between-group differences in the association between cerebellar volume and HADS-SIS score was conducted, with the aim of clarifying as far as possible, whether the potential associations found in the HD data were disease-related (despite no significant-between group difference). This was done by fitting a regression model for volume on all the data and including HADS-SIS, group and their interaction as explanatory variables, with the usual adjustments for age, gender, study site and alcohol intake history. Pronate/supinate-hand task performance and gait (also associated with HD cerebellar volume) could not be included as outcomes in this post hoc analysis due to floor-effects in the control group's scores.

Due to the limited sample size it was decided not to correct for multiple comparisons in the results. P-values were therefore interpreted appropriately, acknowledging the increased risk of false-positive results.

13.3 Results

Participant demographics are detailed in Table 13-1. Controls were on average, older than the HD group. Similar proportions of control and HD participants reported previous or current alcohol abuse (33% and 27% respectively). No participants reported use of gabaergic medications at the time of scanning.

Table 13-1. Cerebellum imaging investigation participant demographics and characteristics.		
	Controls	HD
n = 34	12	22
Age; years (mean (SD) range)	53.8 (6.4) 45-65	48.2 (13.2) 23-67
Gender (M/F)	3/9	9/13
Site (Verio/Tim Trio)	7/5	13/9
CAG (mean (SD))	N/A	43.9 (4.4)
TMS (mean (SD))^a	1.25 (1.54)	18.23 (7.53)
TFC (mean (SD))^b	13 (0)	11.68 (2.03)
Disease burden^c	N/A	353.5 (77.6)
Education (ISCED) (mean (SD) range)^d	4 (1.35) 2-6	3.5 (1.50) 2-6
Alcohol intake history^e	Never abused (n=8) Previous (n=1) Current abuse (n=3)	Never abused (n=16) Previous (n=1) Current abuse (n=5)
Alcohol units per week in previous/current abusers (mean (SD))	20 (5.57)	16 (8.88)
Presence of gabaergic medication^e	None	None
^a TMS and ^b TFC from the UHDRS (1996). ^c Disease burden calculated as: (CAG – 35.5) x current age (Penney et al. 1997). ^d ISCED = International Standard Classification of Education. ^e Alcohol abuse history and usage of gabaergic medications are included as these factors may affect cerebellar structure and functioning (Manto 2012).		

Significant or borderline significant between-group differences were seen in all diffusion metrics (except WM FA) within both the cerebellar GM and WM (Table 13-2). In the HD group diffusion metrics were significantly increased compared with controls and GM FA was significantly reduced. At the voxel-level, there was evidence of WM volume reductions (but no GM differences) in the HD group compared with controls, focused within the paravermis region (Figure 13-2).

Table 13-2. Summary statistics and adjusted between-group differences in cerebellar imaging metrics.				
Measure	Controls (n=12)	HD (n=22)	Adj. Between-Group Difference	Adj. P-value
	Mean (SD)		(HD – Controls; 95% CI)	
Volume/TIV	0.085 (0.009)	0.083 (0.008)	-0.0022 (-0.0078, 0.0034)	0.448
WM FA	0.496 (0.021)	0.490 (0.034)	-0.0021 (-0.0186, 0.0144)	0.801
WM MD (mm ² /s x10 ⁻³)	0.6358 (0.013)	0.6559 (0.037)	0.0226 (0.0061, 0.0391)	0.007
WM AD (mm ² /s x10 ⁻³)	1.0127 (0.037)	1.0398 (0.061)	0.0338 (0.0042, 0.0634)	0.025
WM RD (mm ² /s x10 ⁻³)	0.4473 (0.010)	0.464 (0.035)	0.0154 (-0.0002, 0.031)	0.053
GM FA	0.244 (0.013)	0.228 (0.013)	-0.0185 (-0.0264, -0.0107)	<0.001
GM MD (mm ² /s x10 ⁻³)	0.7618 (0.032)	0.7929 (0.079)	0.0477 (0.0123, 0.0831)	0.008
GM AD (mm ² /s x10 ⁻³)	0.9544 (0.039)	0.9771 (0.090)	0.0403 (0.0006, 0.08)	0.047
GM RD (mm ² /s x10 ⁻³)	0.6655 (0.030)	0.7008 (0.075)	0.0513 (0.0178, 0.0848)	0.003
Between-group differences assessed using generalised least squares regression adjusted for age, gender, study site and alcohol intake history.				

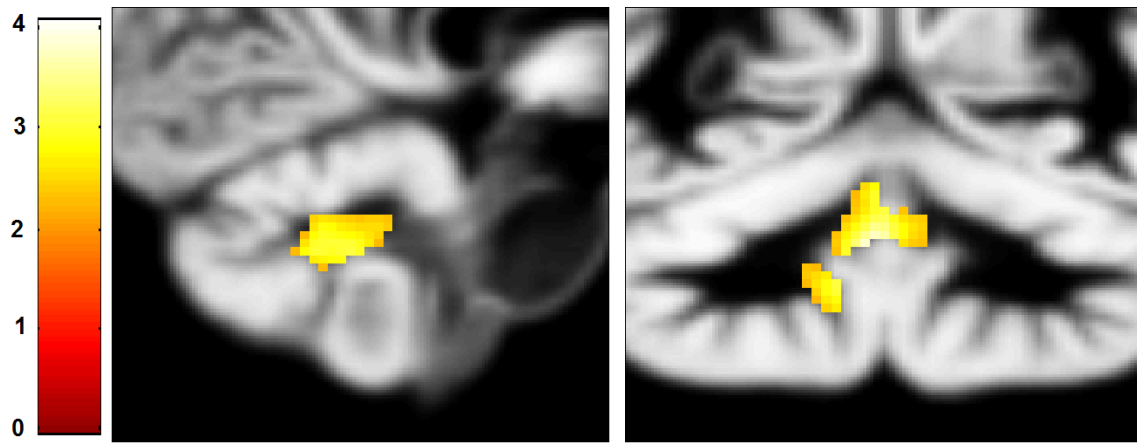


Figure 13-2. A statistical parametric map showing t scores of significant ($p < 0.05$) differences between controls and early HD patients in cerebellar WM volume, adjusted for age, gender, study site, TIV and alcohol intake history. Results are adjusted for multiple comparisons using FDR correction ($p < 0.05$).

Associations between cerebellar imaging metrics and scores on the pre-defined clinical measures are reported in **Error! Reference source not found.** Statistically, the strongest association was found between reduced cerebellar volume and impaired gait. Reduced volume was also significantly associated with psychiatric morbidity, as measured by the HADS-SIS, and associated with borderline significance with impaired pronate/supinate-hand task performance. Increased WM RD was significantly associated with impaired finger tapping. Decreased WM FA and increased diffusion metrics also showed borderline significant associations with one or more of increased TMS, impaired saccade initiation, pronate/supinate-hand task performance and tandem walk. These associations are plotted in Figure 13-3.

Table 13-3. Imaging associations with clinical scores in the HD group (n=22) – association coefficient (95% CI) p-value.

Measures	TMS	Saccade initiation	Pronate/supinate-hand task	Tandem Walk	Gait	Retropulsion-Pull Test	Finger Tapping	HADS-SIS
Volume /TIV	-37.68 (-91.67 , 16.31) p=0.158	-145.05 (-460.93, 170.82) p=0.343	-340.72 (-719.08 , 37.64) p=0.074	-227.79 (-706.1 , 250.52) p=0.326	-744.94 (-1264.56,-225.32) p=0.008	532.82 (-461.44, 1527.09) p=0.271	-93.39 (-530.63 , 343.86) p=0.655	-38.01 (-72.21 , -3.81) p=0.032
WM FA	186.22 (-41.02 , 413.46) p=0.101	1176.32 (-69.88 , 2422.52) p=0.063	501.9 (-1297.16,2300.95) p=0.561	1060.09 (-989.79, 3109.97) p=0.288	1976.49 (-660.9 , 4613.88) p=0.131	337.2 (-4127.03,4801.43) p=0.874	1502.14 (-205.96, 3210.23) p=0.080	67.56 (-101.19 , 236.31) p=0.407
WM MD (mm ² /s)	-0.163 (-0.437 , 0.110) p>0.05	-0.713 (-2.13 , 0.708) p>0.302	-0.771 (-3.11 , 1.57) p>0.05	-0.654 (-2.86 , 1.55) p>0.537	-1.17 (-4.09 , 1.76) p=0.409	1.83 (-2.75, 6.42) p=0.407	-1.3 (-3.16 , 0.563) p=0.158	-0.0257 (-0.207 , 0.155) p=0.766
WM RD (mm ² /s)	-0.227 (-0.463,0.00778) p=0.057	-1.24 (-2.59 , 0.113) p>0.05	-0.890 (-2.77 , 0.992) p=0.329	-1.03 (-3.24 , 1.18) p=0.336	-1.99 (-4.83 , 0.860) p=0.158	0.804 (-3.95 , 5.56) p=0.724	-1.81 (-3.57 , -0.0417) p=0.045	-0.0568 (-0.239 , 0.125) p=0.516
WM AD (mm ² /s)	-0.0349 (-0.615 , 0.546) p>0.05	0.334 (-2.54 , 3.21) p>0.05	-0.533 (-3.85 , 2.78) p>0.05	0.0959 (-4.61 , 4.8) p>0.05	0.472 (-8.28 , 9.22) p>0.05	3.9 (-3.05 , 10.85) p=0.251	-0.275 (-3.86 , 3.31) p>0.05	0.0364 (-0.316 , 0.388) p>0.05
GM FA	-55.89 (-131.86 , 20.08) p=0.138	-329.83 (-754.38 , 94.71) p=0.118	-279.82 (-858.11 , 298.47) p=0.319	-344.03 (-1018.72, 330.65) p=0.294	-640.05 (-1509.4 , 229.3) p=0.137	-699.68 (-2117.38 , 718.02) p=0.309	200.43 (-413.67 , 814.52) p=0.497	17.36 (-38.64 , 73.37) p=0.519
GM MD (mm ² /s)	0.237 (-0.266 , 0.739) p=0.332	0.928 (-1.95 , 3.81) p=0.503	1.16 (-2.58 , 4.89) p=0.520	3.42 (-0.596 , 7.44) p=0.090	3.29 (-2.37 , 8.95) p=0.234	3.29 (-5.84 , 12.42) p=0.454	-1.01 (-4.93 , 2.9) p=0.589	0.0488 (-0.310 , 0.408) p=0.776
GM RD (mm ² /s)	0.255 (-0.224 , 0.734) p=0.274	1.05 (-1.7 , 3.8) p=0.430	1.23 (-2.34 , 4.81) p=0.473	3.39 (-0.447 , 7.23) p=0.079	3.5 (-1.88 , 8.87) p=0.186	3.26 (-5.5 , 12.02) p=0.440	-0.965 (-4.73 , 2.8) p=0.593	0.0463 (-0.299 , 0.391) p=0.779
GM AD (mm ² /s)	0.200 (-0.355 , 0.754) p=0.455	0.690 (-2.47 , 3.85) p=0.649	0.999 (-3.09 , 5.09) p=0.611	3.49 (-0.949 , 7.93) p=0.115	2.88 (-3.4 , 9.16) p=0.344	3.35 (-6.61 , 13.32) p=0.484	-1.11 (-5.37 , 3.15) p=0.587	0.0537 (-0.337 , 0.445) p=0.774

Associations modelled using ordinary least squares regression, adjusted for age, gender, study site and alcohol intake history. All coefficients and CIs are presented as $\times 10^{-5}$. If models showed signs of deviation from the normality assumption, bias-corrected bootstrap CIs were calculated based on 2000 replications, and p-values reported as < or > 0.05. Dark grey highlights significant associations (p<0.05), light grey highlights associations of borderline significance (p<0.08).

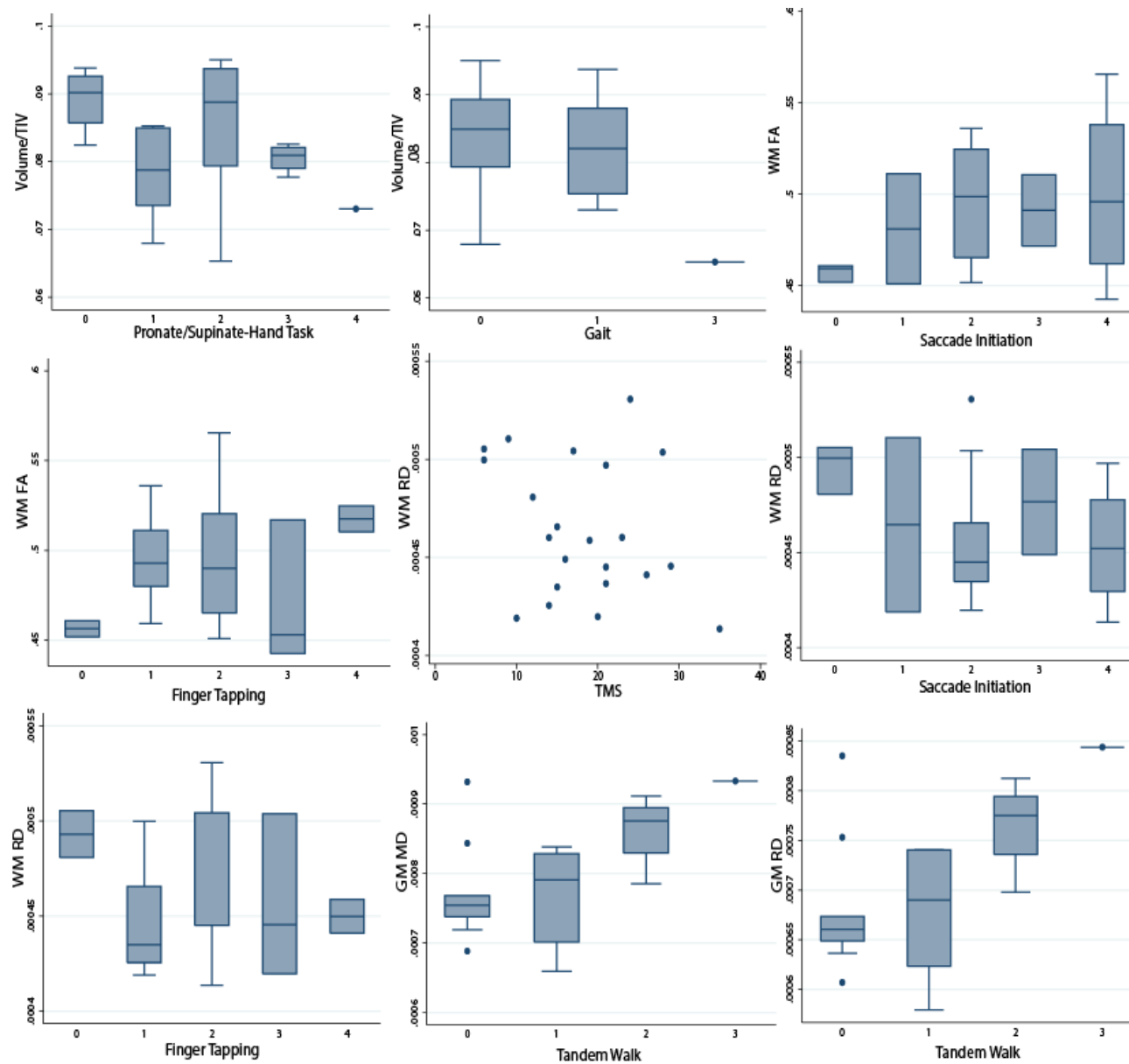


Figure 13-3. Plots of raw HD group data for the significant and borderline associations between clinical scores and imaging metrics.

The association between increased psychiatric morbidity with decreased cerebellar volume appeared to also be present in the control group with a similar estimate (although not statistically significant) as when looking solely at the HD participants (Figure 13-4). A model allowing for an interaction between the effect of HADS-SIS on volume, and study group, gave no evidence of different slopes between HD and control groups ($p>0.05$).

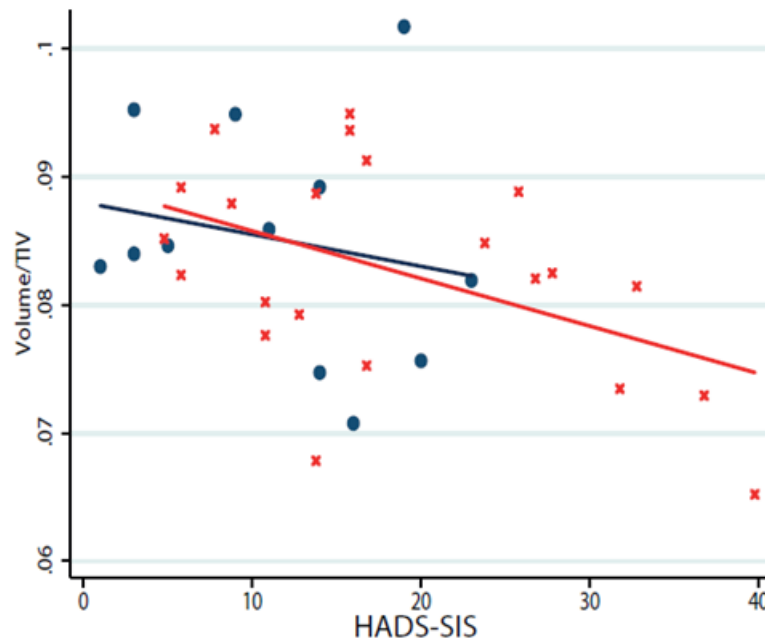


Figure 13-4. Associations between cerebellar volume and HADS-SIS plotted for both the HD (red (x)) and control (blue (•)) groups with the unadjusted regression lines of best fit.

13.4 Discussion

This exploratory investigation of macro- and micro-structural neuroimaging changes in the cerebellum in early-stage HD detected, for the first time, increased diffusion and decreased FA accompanied by reduced volume localised to the paravermal region. Of the hypothesized potential clinical-cerebellar associations, impaired TMS, saccade initiation, pronate/supinate-hand task performance, tandem walk, gait and HADS-SIS score were significantly associated, or of borderline significance, with reduced cerebellar volume and/or increased GM diffusion metrics. Overall these findings suggest a potential role of cerebellar pathology in the HD motor and psychiatric phenotypes.

The reduced volume in the paravermis, the location of the deep cerebellar nuclei, of the HD group compared with controls is consistent with previous autopsy findings (Jeste et al. 1984; Rodda 1981; Rub et al. 2013). This is most likely due to a combination of cell death, cell body shrinkage and other neuronal abnormalities. The deep cerebellar nuclei are not visible on T1-weighted scans therefore optimised scan acquisition parameters would be required to more fully investigate these changes. The sensitivity of VBM to differences in the cerebellum of HD patients in light of an absence of between-group differences

apparent with manual delineation may be due to inherent properties of the two techniques; voxel-wise group comparisons may be more sensitive to volumetric changes within the cerebellum (particularly the WM) because the complex cortical morphology complicates manual border delineation.

Diffusion abnormalities were seen throughout the cerebellum in all four diffusion measures (FA, MD, AD and RD) in either the WM, GM or both. Cerebellar WM in the HD group showed significantly increased MD, AD and a borderline significant increase in RD. This can be interpreted as an increase in the average spacing between membrane layers and 'leakage' out of WM tracts, potentially due to demyelination (Song et al. 2003). The cerebellar GM showed significantly decreased FA and increased MD, AD and RD within the HD group compared with controls. GM metrics are more difficult to interpret as GM is composed of layered, rather than fibrous, tissue. The interpretation here is limited to the conclusion that a general increase in diffusion in all directions within the GM and a decrease in anisotropy implies a pathological change within the micro-structure and composition of the GM tissue, or surrounding glial cells, resulting in abnormal and disorganised water diffusion.

The relationship between the cerebellum and motor impairment is not surprising as the cerebellum is traditionally seen as a motor control centre in the brain (D'Angelo 2011). More recent findings however have implicated the cerebellum in much more varied and higher-level cognitive functions (Allen & Courchesne 1998). Additionally cerebellar dysfunction has been related to psychiatric disorders such as bipolar, psychosis and depression (Kutty & Prendes 1981;Lauterbach 1996;Starkstein et al. 1988), adding weight to the current finding of an association between cerebellar volume and psychiatric morbidity.

Despite observing no significant between-group difference in whole cerebellar volume, assessed using manual delineation, two statistically significant associations with gait and HADS-SIS score were found within the HD group. A previous study in HD found the same result for associations with TFC and disease duration (Rosas et al. 2003). These results cannot necessarily be attributed to disease as no significant volumetric disease-related changes were found. Further analysis of the HADS-SIS data found no evidence that this association differed within the control group. Previous research in a healthy cohort found that individual differences in regional cerebellar volume were associated with sensorimotor and cognitive task performance (Bernard and Seidler 2013). Additional findings have suggested a compensatory function of the cerebellum in disease: a PET study in preHD showed hypermetabolism of the thalamus and cerebellum, suggestive of compensatory function (Feigin et al. 2007); also, in other movement disorders (e.g. Parkinson's Disease) the cerebellum is considered a potential compensatory region (Wu & Hallett 2013). Together this evidence suggests that larger cerebellar volumes may provide a protective functional capacity which could attenuate symptoms in HD, although at this stage this is a speculative hypothesis, requiring validation.

In this well-characterised, homogeneous early-stage HD cohort, clinical signs and cerebellar pathology are mild and therefore this study, although of similar size to previous studies of the cerebellum in HD, was limited in power. It should also be noted that multiple statistical tests were conducted with no adjustment for multiple comparisons. Future work would be enhanced by investigating larger cohorts which might give more precise estimation and add strength to the statistical associations detected in this cohort should they not be chance findings. Due to limited cerebellar scan coverage in a significant proportion of the PADDINGTON cohort only a subset was useable for this study. This selection based on coverage could have biased the sample towards those with smaller head and cerebellar sizes, potentially reducing the likelihood of detecting between-group volumetric differences. It should be noted that analysis of cerebellar volume as a proportion of TIV (to adjust for inter-subject head-size) assumes a linear relationship between the two which may not be the case. This is however the standard way to report regional volumetric data. A further limitation is that the control group was notably older than the HD group. However, the direction of the hypothesized effect (age-related degeneration) means that this age difference is likely to have made significant disease effects harder to detect, rather than introduce a disease bias.

This was an exploratory analysis of volumetric and diffusion cerebellar imaging metrics and their clinical associations in HD. Associations were found between reduced cerebellar volume and pathological diffusion and several motor and psychiatric symptoms of stage I HD. Future work would be enhanced by investigating larger cohorts, using imaging sequences more tailored to the cerebellum and focusing on diffusion and functional, over volumetric, cerebellar changes in HD.

14 Emotion Recognition in HD: An Exploratory Imaging Investigation

14.1 Background

The functional imaging literature typically advocates the idea of discrete neural substrates underlying the processing of specific emotions, for example: the insula and basal ganglia are thought to be associated with disgust processing (Adolphs 2002; Sprengelmeyer et al. 1998), the cingulate gyrus, left medial temporal gyrus and the ventral striatum with anger, and the amygdala with fear (Adolphs 2002). Additionally there are regions known to be associated with facial processing (the fusiform gyrus (Kanwisher et al. 1997)) and the perception of expressive movements (the superior temporal gyrus (STG) (Puce et al. 1998)), both of which are necessary for the recognition of emotion from facial expressions and body language.

Emotion recognition performance in HD has been linked with limbic GM structures (Ille et al. 2011a; Kipps et al. 2007) and regions responsible for executive function, memory, visual and mental imagery (Henley et al. 2008; Ille et al. 2011a; Scahill et al. 2011). WM volumes within the left external capsule and right parietal area have also been linked with HD-related emotion recognition deficits (Scahill et al. 2011). Contrary to the idea of discrete emotion-specific neural substrates, Henley et al. (Henley et al. 2008) propose a generic fronto-subcortical network in the pathogenesis of emotion recognition deficits in HD.

The novel test of emotion recognition, described in Chapter 10, detected deficits in disgust, anger and fear recognition in an early-stage HD cohort; potentially indicative of a common, damaged cognitive process or system underlying these three emotions. There were however significant differences in the emotion-specific pattern of the impairment when tested on different stimulus modalities (photos, vocal expressions and film clips). For example, vocal expressions of anger were recognised by the HD group at a significantly closer level to the control group performance than anger recognition from dynamic film clips. In addition, the impairment in fear recognition was estimated to be greater than that for anger in the vocal stimuli, whereas the converse was true for film stimuli. These modality-specific variations in the impairment are suggestive of slightly divergent underlying cognitive processes. Much of what is known from the literature about the neural correlates of emotion recognition is derived from functional imaging studies based on recognition from static photo stimuli (Adolphs 2002). This modality has been shown not to be fully representative of emotion recognition ability from other stimulus modalities.

This exploratory imaging investigation aimed to test whether the HD group's performance levels on emotion-specific recognition and/or performance within the three stimulus modalities (photos, vocal expressions and film clips) can be explained by, or is associated with, dissociable underlying GM or WM pathology or a generic emotion network. This is the first study in HD to explore associations between the micro-structure of cerebral WM, using diffusion imaging, and emotion recognition and to investigate the neural associates of emotion recognition via different stimulus modalities.

14.2 Methods

14.2.1 Image Acquisition

All 15 HD participants who took part in the emotion recognition task (Section 10.3) were scanned during their visit at the London site of the PADDINGTON study. 3T MRI data (T1- and diffusion-weighted) were acquired based on protocols previously described (Section 3.4). Data were pseudoanonymised and archived on a secure web-portal.

14.2.2 Image Analysis

QC was performed on all scans, checking for artefacts (e.g. movement, intensity) and sufficient tissue contrast for analysis. All scans were deemed to be of good quality.

For the macro-structural analyses, ROIs were selected based on the current emotion recognition literature (Adolphs 2002). The caudate and TIV were manually delineated in MIDAS software (Freeborough et al. 1997), using protocols previously tested and published (Hobbs et al. 2009; Whitwell et al. 2001). Putamen, amygdala and globus pallidus volumes were generated using FIRST from FSL (version 5 (Patenaude et al. 2011)) with an additional brain extraction step and the FAST boundary correction option. Volumes were measured in ml and converted to a percentage of TIV. Cortical thickness measures (fusiform, STG and insula) were obtained using FreeSurfer version 5.1.0 (Fischl et al. 1999) and are reported in mm. All segmentations were visually inspected for quality.

The micro-structural diffusion analysis was run using DTI-TK (Zhang et al. 2007). A whole-brain study-specific group template was created using high-dimensional spatial normalisation of the tensor image from each participant. Native-space images and the JHU-ICBM atlas (Wakana et al. 2007) were warped to this template and FA maps generated for all participants. All atlas regions were included in the analysis, with the exception of the middle cerebral peduncle and tapetum (deemed to be too small for reliable analyses). To remove voxels contaminated with partial volumes the JHU regions were thresholded to only include voxels with group template $FA > 0.2$. Average FA readings were calculated from each of the 45 regions.

14.2.3 Statistical Analysis

Due to the relatively small HD sample size ($n=15$) associations between imaging measures (volume/thickness and regional FA) and emotion recognition score were assessed for each emotion across all modalities, or for each stimulus modality across all emotions. Ordinary least squares regression was used, fitting a separate model for each emotion (or modality). These imaging analyses were adjusted for age, gender, BFRT and motor response time. Education level is not known to be associated with disease progression (manifested as brain pathology) therefore this was not considered necessary as a covariate with the imaging associations. The relative homogeneity of the HD cohort meant that it was not considered

necessary to adjust for disease burden. If examination of model residuals suggested non-normality, bias-corrected and accelerated bootstrap CIs were calculated for the group difference and accuracy was limited to > 0.05 or < 0.01 . As advocated by Rothman (Rothman 1990) no adjustment was made for multiple comparisons due to the small sample size and the exploratory nature of the regional associations, therefore results have been interpreted accordingly.

14.3 Results

14.3.1 Macro-Structural Associations

For all regions, reduced volume or thickness was related to poorer performance on the task overall ('Total Errors' /90 (Table 14-1 – left column)). This negative association reached statistical significance with caudate volume (as a percentage of TIV; $p=0.018$). Caudate volume showed relatively consistent estimates across all stimulus modalities, with the expected increase in errors per 0.1% absolute decrease in caudate volume (as a %TIV) being -4.95 (-11.52, 1.62), -4.90 (-9.54, -0.26) and -6.5 (-11.35, -1.65) for photo, vocal and film errors respectively. Cortical thickness estimates within the fusiform and superior temporal gyri were found to be significantly associated ($p<0.05$) with the number of vocal emotion cue errors. This association was of borderline significance in the insula also ($p=0.058$).

	Total Errors		Photo Errors		Vocal Errors		Film Errors	
	Estimate (95% CI)	P- Value	Estimate (95% CI)	P- Value	Estimate (95% CI)	P- Value	Estimate (95% CI)	P- Value
Caudate (%TIV)	-16.35 (-28.98, -3.71)	0.018	-4.95 (-11.52, 1.62)	0.121	-4.90 (-9.54, -0.26)	0.041	-6.50 (-11.35, -1.65)	0.015
Putamen (%TIV)	-9.01 (-28.83, 10.82)	0.325	-2.32 (-10.99, 6.36)	0.555	-3.40 (-9.88, 3.08)	0.261	-3.29 (-11.11, 4.53)	0.360
Amygdala (%TIV)	-4.41 (-29.43, 20.60)	0.695	-4.63 (-14.55, 5.29)	0.313	1.05 (-7.34, 9.43)	0.781	-0.82 (-10.69, 9.04)	0.852
Globus Pallidus (%TIV)	-11.17 (-42.89, 20.54)	0.440	-4.04 (-17.50, 9.43)	0.509	-5.76 (-15.71, 4.18)	0.218	-1.37 (-14.23, 11.49)	0.812
Fusiform Gyrus (mm)	-15.2 (-31.55, 1.14)	0.064	-5.61 (-12.91, 1.69)	0.114	-5.52 (-10.67, -0.38)	0.038	-4.07 (-11.38, 3.24)	0.235
STG (mm)	-17.06 (-42.75, 8.63)	0.164	-5.21 (-16.72, 6.30)	0.327	-8.57 (-15.35, -1.78)	0.020	-3.29 (-14.40, 7.82)	0.514
Insula (mm)	-9.91 (-34.57, 14.75)	0.381	-0.98 (-11.85, 9.88)	0.840	-6.53 (-13.35, 0.28)	0.058	-2.39 (-12.35, 7.57)	0.595

For volumes, the estimates represent the expected change in number of errors in the stimulus modality per 0.1% absolute increase in volume as a percentage of TIV. For thicknesses, the estimates represent the expected change in number of errors in the stimulus modality per mm increase in thickness. All estimates are adjusted for age, gender, BFRT score and motor response time. Dark grey highlights significant associations ($p<0.05$), light grey highlights associations of borderline significance ($p<0.08$).

Table 14-2 presents associations between regional brain volumes or thickness and the number of errors made within each emotion (combined across stimulus modalities). When looking at specific emotions caudate volume was significantly associated ($p < 0.05$) with anger and surprise recognition; although caudate volume associations with other emotions gave an overall suggestion that more errors was associated with decreased volume (all were negative). The fusiform and STG thicknesses were significantly associated with the number of errors in fear recognition and the insula was significantly associated with the number of sadness recognition errors. There were also borderline significant ($p < 0.08$) associations between putamen volume and happiness recognition and fusiform thickness and anger recognition. All of these statistically significant results, and results of borderline significance, suggested that a decreased volume or thickness in each region was related to an increase in number of errors, and although the different regions did not all have significant associations with the same emotions, all of the estimates and p-values were broadly similar.

	Anger Errors	Disgust Errors	Fear Errors	Happiness Errors	Sadness Errors	Surprise Errors
Caudate (%TIV)	-4.78 (-9.12 , -0.44) $p=0.034$	-1.73 (-5.19 , 1.73) $p=0.287$	-2.27 (-5.3 , 0.77) $p=0.125$	-1.11 (-2.91 , 0.7) $p=0.199$	-0.82 (-4.43 , 2.8) $p=0.622$	-2.52 (-4.31 , -0.74) $p=0.011$
Putamen (%TIV)	-2.94 (-7.48 , 1.6) $p=0.177$	-0.44 (-3.73 , 2.86) $p=0.772$	-0.6 (-3.68 , 2.49) $p=0.672$	-1.39 (-2.83 , 0.04) $p=0.056$	-0.6 (-3.85 , 2.65) $p=0.686$	-1.38 (-3.47 , 0.71) $p=0.168$
Amygdala (%TIV)	-1.92 (-13.38 , 9.54) $p=0.714$	-0.28 (-7.85 , 7.28) $p=0.935$	-4.74 (-10.9 , 1.42) $p=0.116$	-0.97 (-4.97 , 3.03) $p=0.597$	-5.07 (-11.52 , 1.38) $p=0.109$	-0.2 (-5.54 , 5.13) $p=0.933$
Globus Pallidus (%TIV)	2.25 (-6.98 , 11.49) $p=0.594$	-2.21 (-8.13 , 3.72) $p=0.421$	-2.04 (-7.62 , 3.55) $p=0.431$	0.32 (-2.98 , 3.62) $p=0.83$	0.56 (-5.53 , 6.64) $p=0.841$	-0.3 (-4.63 , 4.03) $p=0.879$
Fusiform gyrus (mm)	-5.66 (-12.07 , 0.74) $p=0.077$	-2.95 (-7.48 , 1.58) $p=0.175$	-3.99 (-7.66 , -0.32) $p=0.036$	1.27 (-1.27 , 3.8) $p=0.288$	-3.03 (-7.48 , 1.42) $p=0.158$	-0.98 (-4.46 , 2.5) $p=0.541$
STG (mm)	-3.03 (-13.48 , 7.43) $p=0.529$	-2.07 (-8.9 , 4.77) $p=0.511$	-5.59 (-10.67 , -0.51) $p=0.034$	0.95 (-2.75 , 4.65) $p=0.575$	-5.02 (-10.85 , 0.8) $p=0.083$	-0.03 (-4.97 , 4.92) $p=0.991$
Insula (mm)	-2.29 (-11.94 , 7.35) $p=0.603$	0.26 (-6.15 , 6.68) $p=0.928$	-2.69 (-8.38 , 3) $p=0.312$	-2.15 (-5.2 , 0.89) $p=0.144$	-5.33 (-10.26 , -0.4) $p=0.037$	-0.8 (-5.29 , 3.68) $p=0.696$
Estimates represent the expected change in number of errors in the emotion per 0.1% absolute increase in volume as a percentage of TIV. For thicknesses, the estimates represent the expected change in number of errors in the emotion per mm increase in thickness. Dark ($p < 0.05$) and light ($p < 0.08$).grey highlights borderline and significant associations.						

14.3.2 Micro-Structural Associations

Reduced FA within the right superior fronto-occipital fasciculus (SFOF) was the only diffusion measure to be significantly associated with overall emotion recognition performance combined across all modalities ('Total Errors' /90 (Table 14-3 – left column); $p=0.032$). When the three stimulus modalities were analysed separately more widespread FA reductions were found to significantly associate with poorer emotion recognition performance, although no clear patterns of associations were identifiable.

The results of regression analysis between regional FA averages and emotion-specific errors (across all three stimulus modalities) are shown in Table 14-4. For the significant associations poorer performance was associated with decreased FA in all but the cingulum-happiness association. These emotion-FA associations highlight six WM regions where the most consistent relationships were observed: the cerebral peduncles, internal capsule; anterior corona radiata, SFOF, cingulum and CC.

Table 14-3. Micro-structural FA associations, across all JHU-atlas regions tested, with emotion recognition errors within stimulus modalities. Estimates (95% CI) and p-values are adjusted for age, gender, BFRT score and motor response time.

	Total Errors	Photo Errors	Vocal Errors	Film Errors
Pontine crossing tract	0.395 (-24.884 , 25.675) p=0.973	-0.249 (-7.162 , 6.663) p=0.937	3.241 (-1.875 , 8.356) p=0.186	0.894 (-6.311 , 8.099) p=0.785
Corpus callosum - Genu	-11.188 (-45.515 , 23.139) p=0.480	-6.670 (-14.927 , 1.587) p=0.101	-3.411 (-10.908 , 4.087) p=0.33	-5.631 (-14.815 , 3.552) p=0.199
Corpus callosum - Body	-14.384 (-41.914 , 13.147) p=0.268	-4.665 (-11.954 , 2.623) p=0.182	-2.041 (-8.495 , 4.413) p=0.492	-4.534 (-12.283 , 3.214) p=0.218
Corpus callosum - Splenium	-3.685 (-41.631 , 34.261) p=0.831	-6.414 (-15.629 , 2.8) p=0.150	0.888 (-7.617 , 9.394) p=0.819	-7.216 (-16.65 , 2.219) p=0.118
Fornix	-7.965 (-31.114 , 15.185) p=0.456	-4.448 (-10.064 , 1.169) p=0.107	-4.465 (-8.639 , -0.29) p=0.039	-3.675 (-9.935 , 2.585) p=0.217
Corticospinal tract (R)	-24.411 (-54.090 , 5.268) p=0.096	-3.878 (-12.972 , 5.216) p=0.360	-6.553 (-12.629 , -0.478) p=0.037	-4.750 (-14.083 , 4.583) p=0.279
Corticospinal tract (L)	-12.320 (-41.381 , 16.740) p=0.363	-6.971 (-13.452 , -0.489) p=0.038	-4.309 (-10.33 , 1.711) p=0.14	-6.217 (-13.585 , 1.151) p=0.089
Medial lemniscus (R)	-13.812 (-49.635 , 22.011) p=0.406	-3.020 (-12.967 , 6.926) p=0.509	-2.236 (-10.430 , 5.958) p=0.552	-8.360 (-16.98 , 0.259) p=0.056
Medial lemniscus (L)	4.378 (-25.395 , 34.152) p=0.747	-6.510 (-13.07 , 0.051) p=0.051	-1.259 (-7.909 , 5.391) p=0.678	-7.831 (-14.048 , -1.613) p=0.019
Inferior cerebellar peduncle (R)	-4.194 (-27.552 , 19.163) p=0.694	-2.963 (-9.01 , 3.085) p=0.297	-0.304 (-5.585 , 4.976) p=0.899	-4.679 (-10.430 , 1.072) p=0.099
Inferior cerebellar peduncle (L)	6.313 (-18.393 , 31.020) p=0.577	-4.833 (-10.671 , 1.006) p=0.094	1.536 (-3.985 , 7.058) p=0.545	-4.969 (-11.12 , 1.182) p=0.101
Superior cerebellar peduncle (R)	-4.532 (-32.958 , 23.893) p=0.727	-5.432 (-12.106 , 1.243) p=0.099	-1.948 (-8.198 , 4.302) p=0.499	-5.315 (-12.463 , 1.833) p=0.127
Superior cerebellar peduncle (L)	-3.725 (-29.958 , 22.508) p=0.755	-2.915 (-9.788 , 3.959) p=0.362	-0.571 (-6.471 , 5.329) p=0.832	-3.436 (-10.529 , 3.658) p=0.302
Cerebral peduncle (R)	-6.298 (-70.166 , 57.569) p=0.828	-9.284 (-25.341 , 6.774) p=0.223	-7.917 (-20.978 , 5.144) p=0.204	-11.524 (-27.665 , 4.618) p=0.141
Cerebral peduncle (L)	-8.958 (-41.125 , 23.21) p=0.544	-7.132 (-14.337 , 0.072) p=0.052	-4.230 (-10.875 , 2.414) p=0.184	-9.171 (-15.55 , -2.791) p=0.01
Anterior limb internal capsule (R)	-12.933 (-51.534 , 25.668) p=0.468	-8.295 (-17.21 , 0.619) p=0.065	-7.132 (-14.259 , -0.006) p=0.05	-8.335 (-17.842 , 1.171) p=0.079
Anterior limb internal capsule (L)	-20.121 (-69.945 , 29.703) p=0.385	-7.887 (-20.832 , 5.057) p=0.201	-10.552 (-19.101 , -2.003) p=0.021	-8.105 (-21.703 , 5.492) p=0.21
Posterior limb internal capsule (R)	-24.548 (-68.834 , 19.738) p=0.241	-4.860 (-17.467 , 7.747) p=0.406	-9.060 (-17.377 , -0.743) p=0.036	-12.352 (-22.452 , -2.252) p=0.022
Posterior limb internal capsule (L)	-0.500 (-59.074 , 58.074) p=0.985	-13.296 (-25.793 , -0.8) p=0.039	-9.802 (-20.66 , 1.052) p=0.076	-15.964 (-27.635 , -4.292) p=0.013
Retrolenticular internal capsule (R)	3.373 (-45.146 , 51.893) p=0.878	-7.612 (-19.598 , 4.374) p=0.185	-3.902 (-14.392 , 6.588) p=0.422	-10.569 (-21.966 , 0.828) p=0.065
Retrolenticular internal capsule (L)	3.668 (-23.057 , 30.392) p=0.763	-0.546 (-7.884 , 6.791) p=0.870	0.045 (-5.979 , 6.07) p=0.987	-2.563 (-10.007 , 4.881) p=0.456
Anterior corona radiata (R)	-4.405 (-45.338 , 36.527) p=0.813	-10.132 (-18.366 , -1.898) p=0.021	-4.719 (-13.212 , 3.775) p=0.24	-7.050 (-17.534 , 3.434) p=0.163
Anterior corona radiata (L)	4.302 (-45.215 , 53.818) p=0.849	-8.097 (-20.219 , 4.025) p=0.165	0.011 (-11.117 , 11.138) p=0.998	-3.687 (-17.616 , 10.242) p=0.564
Superior corona radiata (R)	21.811 (-15.931 , 59.553) p=0.224	-8.642 (-17.826 , 0.541) p=0.062	-2.265 (-11.337 , 6.808) p=0.586	-7.038 (-17.56 , 3.483) p=0.165
Superior corona radiata (L)	24.326 (-18.968 , 67.62) p=0.236	-9.636 (-20.248 , 0.975) p=0.070	-2.063 (-12.491 , 8.365) p=0.665	-7.554 (-19.748 , 4.64) p=0.195

	Total Errors	Photo Errors	Vocal Errors	Film Errors
Posterior corona radiata (R)	0.372 (-30.437, 31.182) p=0.979	-6.781 (-13.48, -0.082) p=0.048	-2.998 (-9.526, 3.531) p=0.326	-6.056 (-13.601, 1.490) p=0.103
Posterior corona radiata (L)	19.644 (-10.605, 49.892) p=0.176	-4.416 (-13.005, 4.173) p=0.275	0.818 (-6.709, 8.345) p=0.811	-2.167 (-11.668, 7.334) p=0.618
Posterior thalamic radiation (R)	17.909 (-21.047, 56.865) p=0.325	-2.973 (-14.026, 8.079) p=0.558	2.620 (-6.412, 11.652) p=0.528	-3.253 (-14.797, 8.291) p=0.54
Posterior thalamic radiation (L)	1.610 (-36.056, 39.277) p=0.925	-1.427 (-11.679, 8.824) p=0.760	-0.156 (-8.606, 8.294) p=0.968	-3.407 (-13.884, 7.07) p=0.481
Sagittal stratum (R)	-9.413 (-52.899, 34.072) p=0.636	-4.315 (-15.919, 7.289) p=0.422	-3.302 (-12.863, 6.26) p=0.455	-5.718 (-17.57, 6.135) p=0.303
Sagittal stratum (L)	-26.823 (-72.508, 18.862) p=0.217	-3.685 (-17.065, 9.695) p=0.549	-7.834 (-17.354, 1.685) p=0.096	-5.055 (-18.839, 8.728) p=0.428
External capsule (R)	-4.044 (-73.916, 65.828) p=0.899	-13.171 (-29.521, 3.178) p=0.102	-7.474 (-22.109, 7.162) p=0.278	-10.504 (-28.89, 7.881) p=0.228
External capsule (L)	1.642 (-59.479, 62.763) p=0.953	-12.793 (-26.451, 0.866) p=0.063	-8.451 (-20.588, 3.687) p=0.15	-12.211 (-27.091, 2.670) p=0.096
Cingulum cingulate (R)	7.230 (-21.939, 36.399) p=0.589	-6.115 (-12.794, 0.565) p=0.068	-0.288 (-6.939, 6.362) p=0.924	-3.722 (-11.739, 4.295) p=0.321
Cingulum cingulate (L)	0.524 (-45.232, 46.279) p=0.98	-13.783 (-20.756, -6.809) p=0.002	-4.722 (-14.345, 4.901) p=0.296	-11.192 (-21.208, -1.176) p=0.032
Cingulum hippocampus (R)	5.138 (-35.041, 45.317) p=0.779	-3.695 (-14.378, 6.989) p=0.454	0.561 (-8.481, 9.602) p=0.892	0.868 (-10.668, 12.404) p=0.869
Cingulum hippocampus (L)	-4.229 (-24.787, 16.329) p=0.653	-3.745 (-8.685, 1.196) p=0.121	-2.082 (-6.475, 2.312) p=0.312	-5.056 (-9.630, -0.482) p=0.034
Fornix cres Stria terminalis (R)	-10.724 (-40.001, 18.553) p=0.429	-2.427 (-10.531, 5.677) p=0.515	-2.962 (-9.396, 3.472) p=0.325	-2.542 (-11.022, 5.938) p=0.515
Fornix cres Stria terminalis (L)	-10.427 (-36.795, 15.94) p=0.394	-4.932 (-11.475, 1.611) p=0.122	-3.855 (-9.297, 1.587) p=0.144	-4.783 (-11.785, 2.218) p=0.157
Superior longitudinal fasciculus (R)	7.972 (-29.804, 45.749) p=0.644	-9.290 (-17.061, -1.519) p=0.024	-4.086 (-12.091, 3.919) p=0.278	-6.248 (-16.132, 3.635) p=0.186
Superior longitudinal fasciculus (L)	9.323 (-30.402, 49.047) p=0.608	-7.296 (-16.861, 2.268) p=0.118	-0.518 (-9.555, 8.52) p=0.9	-3.122 (-14.427, 8.183) p=0.548
SFOF (R)	-25.585 (-48.503, -2.666) p=0.032	0.310 (-7.88, 8.501) p=0.934	-6.080 (-10.99, -1.170) p=0.021	-1.797 (-10.264, 6.671) p=0.643
SFOF (L)	-5.593 (-49.838, 38.653) p=0.781	-12.037 (-20.125, -3.95) p=0.008	-8.019 (-15.942, -0.097) p=0.048	-10.177 (-20.324, -0.030) p=0.049
Uncinate fasciculus (R)	-14.335 (-33.124, 4.454) p=0.118	0.273 (-5.652, 6.198) p=0.919	-1.422 (-6.163, 3.318) p=0.514	-2.252 (-8.220, 3.715) p=0.415
Uncinate fasciculus (L)	-2.284 (-30.232, 25.665) p=0.857	0.004 (-7.655, 7.663) p=0.999	-0.882 (-7.126, 5.362) p=0.757	1.308 (-6.646, 9.262) p=0.718
Estimates represent expected change in number of errors within the stimulus modality per 0.1 unit increase in FA. Dark grey highlights significant associations (p<0.05), light grey highlights associations of borderline significance (p<0.08).				

Table 14-4. Micro-structural FA associations, across all JHU-atlas regions tested, with emotion-specific recognition errors. Estimates (95% CI) and p-values are adjusted for age, gender, BFRT score and motor response time.

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Pontine crossing tract	2.07 (-4.42, 8.57) p=0.488	-1.7 (-5.89, 2.49) p=0.382	2.52 (-1.14, 6.18) p=0.154	0.26 (-2.08, 2.61) p=0.805	0.7 (-3.61, 5.01) p=0.722	0.3 (-2.78, 3.38) p=0.832
Corpus callosum - Genu	-1.35 (-10.63, 7.94) p=0.75	-5.16 (-9.89, -0.44) p=0.035	-3.41 (-8.56, 1.75) p=0.169	1.27 (-1.87, 4.42) p=0.383	-1.92 (-7.81, 3.97) p=0.48	-1.38 (-5.57, 2.81) p=0.476
Corpus callosum - Body	-1.58 (-9.31, 6.14) p=0.654	-4.26 (-8.25, -0.27) p=0.039	-1.88 (-6.49, 2.73) p=0.381	1.19 (-1.42, 3.79) p=0.33	-0.42 (-5.49, 4.65) p=0.856	-1.75 (-5.11, 1.61) p=0.269
Corpus callosum - Splenium	-3.29 (-13.03, 6.46) p=0.465	-5.38 (-10.57, -0.19) p=0.044	-1.09 (-7.23, 5.06) p=0.698	2.26 (-0.84, 5.36) p=0.133	3.26 (-2.8, 9.31) p=0.254	-3.3 (-7.22, 0.62) p=0.089
Fornix	-3.02 (-8.92, 2.87) p=0.276	-2.43 (-6.14, 1.28) p=0.173	-3.77 (-6.44, -1.1) p=0.011	0.58 (-1.6, 2.76) p=0.56	-1.79 (-5.67, 2.09) p=0.324	-1.01 (-3.83, 1.8) p=0.436
Corticospinal tract (R)	-4.57 (-13.13, 3.99) p=0.258	-0.78 (-6.79, 5.24) p=0.777	-3.6 (-8.6, 1.4) p=0.138	-1.14 (-4.27, 1.99) p=0.432	-4.59 (-9.48, 0.3) p=0.063	-0.96 (-5.16, 3.25) p=0.619
Corticospinal tract (L)	-5.03 (-12.15, 2.08) p=0.144	-3.82 (-8.25, 0.61) p=0.083	-2.8 (-7.3, 1.71) p=0.194	-0.1 (-2.94, 2.73) p=0.936	-3.68 (-8.12, 0.76) p=0.093	-1.66 (-5.17, 1.85) p=0.313
Medial lemniscus (R)	-5.71 (-14.58, 3.16) p=0.179	-2.32 (-8.54, 3.9) p=0.421	0.25 (-5.83, 6.32) p=0.929	-1.53 (-4.8, 1.75) p=0.319	1.22 (-5.11, 7.56) p=0.672	-3.14 (-7.03, 0.75) p=0.101
Medial lemniscus (L)	-5.52 (-12.25, 1.21) p=0.096	-3.65 (-8.04, 0.75) p=0.094	-2.68 (-7.12, 1.77) p=0.206	0.97 (-1.72, 3.66) p=0.435	2.2 (-2.66, 7.07) p=0.333	-3.38 (-6, -0.75) p=0.017
Inferior cerebellar peduncle (R)	-2.35 (-8.32, 3.62) p=0.397	-1.67 (-5.55, 2.22) p=0.357	-1.04 (-4.8, 2.72) p=0.546	0.26 (-1.93, 2.44) p=0.795	2.04 (-1.7, 5.78) p=0.248	-2.07 (-4.49, 0.35) p=0.085
Inferior cerebellar peduncle (L)	-2.75 (-9.07, 3.57) p=0.351	-3.41 (-6.93, 0.11) p=0.056	-0.15 (-4.25, 3.95) p=0.937	1.17 (-1, 3.34) p=0.255	1.99 (-2.06, 6.04) p=0.296	-1.98 (-4.67, 0.7) p=0.129
Superior cerebellar peduncle (R)	-2.38 (-9.73, 4.97) p=0.482	-3.4 (-7.64, 0.84) p=0.103	-2.23 (-6.59, 2.12) p=0.275	-0.09 (-2.75, 2.57) p=0.941	0.05 (-4.86, 4.97) p=0.981	-2.24 (-5.3, 0.82) p=0.132
Superior cerebellar peduncle (L)	-1.02 (-7.94, 5.91) p=0.748	-1.13 (-5.62, 3.36) p=0.584	-1.56 (-5.69, 2.58) p=0.417	-0.21 (-2.66, 2.24) p=0.851	0.49 (-4.03, 5) p=0.813	-1.51 (-4.52, 1.5) p=0.285
Cerebral peduncle (R)	-9.23 (-24.65, 6.2) p=0.209	-9.37 (-17.92, -0.82) p=0.035	-3.74 (-13.79, 6.31) p=0.422	2.21 (-3.51, 7.93) p=0.404	0.62 (-10.36, 11.6) p=0.901	-5.02 (-11.86, 1.83) p=0.132
Cerebral peduncle (L)	-7.46 (-14.08, -0.84) p=0.031	-4.45 (-9.05, 0.15) p=0.056	-3.31 (-8.05, 1.43) p=0.148	0.88 (-2.11, 3.86) p=0.524	1.3 (-4.25, 6.86) p=0.609	-3.9 (-6.63, -1.17) p=0.01
Anterior limb internal capsule (R)	-7.15 (-16.18, 1.89) p=0.107	-2.04 (-8.76, 4.68) p=0.51	-5.91 (-10.63, -1.19) p=0.02	-1.16 (-4.76, 2.44) p=0.485	-4.33 (-10.33, 1.68) p=0.138	-2.26 (-6.81, 2.3) p=0.291
Anterior limb internal capsule (L)	-8.04 (-20.39, 4.32) p=0.175	-0.65 (-9.66, 8.36) p=0.874	-7.36 (-13.78, -0.94) p=0.029	-2.03 (-6.62, 2.57) p=0.345	-7.29 (-14.34, -0.24) p=0.044	-1.07 (-7.38, 5.24) p=0.71
Posterior limb internal capsule (R)	-10.41 (-20.37, -0.46) p=0.042	-3.23 (-11.18, 4.72) p=0.382	-3.01 (-10.5, 4.48) p=0.387	-0.4 (-4.85, 4.06) p=0.844	-0.88 (-9.09, 7.33) p=0.813	-4.29 (-9.18, 0.59) p=0.078
Posterior limb internal capsule (L)	-16.64 (-25.7, -7.58) p=0.002	-6.68 (-15.48, 2.13) p=0.121	-5.01 (-13.78, 3.75) p=0.228	0.65 (-4.78, 6.07) p=0.793	-4.02 (-13.6, 5.57) p=0.368	-4.88 (-11.01, 1.26) p=0.106
Retrolenticular internal capsule (R)	-10.9 (-20.76, -1.05) p=0.034	-3.97 (-11.84, 3.89) p=0.282	-1.51 (-9.35, 6.32) p=0.673	0.61 (-3.89, 5.1) p=0.767	-0.1 (-8.44, 8.24) p=0.979	-3.95 (-9.08, 1.18) p=0.115
Retrolenticular internal capsule (L)	-2.97 (-9.71, 3.77) p=0.345	-1.15 (-5.72, 3.42) p=0.584	1.6 (-2.62, 5.81) p=0.414	0.79 (-1.64, 3.21) p=0.483	0.72 (-3.86, 5.3) p=0.729	-0.71 (-3.94, 2.53) p=0.634
Anterior corona radiata (R)	-3.74 (-14.21, 6.74) p=0.441	-6.29 (-11.59, -1) p=0.025	-5.43 (-10.72, -0.13) p=0.046	2.17 (-1.28, 5.62) p=0.188	-3.23 (-9.85, 3.38) p=0.298	-1.55 (-6.43, 3.32) p=0.489

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anterior corona radiata (L)	-1.31 (-14.39 , 11.76) p=0.825	-7.49 (-13.97 , -1.01) p=0.028	-0.95 (-9.01 , 7.1) p=0.795	2.19 (-2.12 , 6.5) p=0.28	0.08 (-8.44 , 8.59) p=0.984	-1.08 (-7.09 , 4.92) p=0.692
Superior corona radiata (R)	-8.07 (-17.09 , 0.94) p=0.074	-4.23 (-10.61 , 2.14) p=0.167	-2.61 (-9.03 , 3.8) p=0.381	2.63 (-0.65 , 5.9) p=0.103	-1.53 (-8.5 , 5.45) p=0.632	-1.3 (-6.23 , 3.64) p=0.567
Superior corona radiata (L)	-8.47 (-19.13 , 2.18) p=0.106	-3.63 (-11.3 , 4.04) p=0.312	-3.7 (-10.83 , 3.44) p=0.272	2.62 (-1.29 , 6.52) p=0.164	-1.23 (-9.24 , 6.79) p=0.737	-1.66 (-7.27 , 3.94) p=0.519
Posterior corona radiata (R)	-6.35 (-12.93 , 0.24) p=0.057	-2.92 (-7.78 , 1.94) p=0.207	-2.55 (-7.19 , 2.1) p=0.246	0.52 (-2.32 , 3.35) p=0.691	-2.1 (-7.14 , 2.95) p=0.372	-1.92 (-5.39 , 1.55) p=0.242
Posterior corona radiata (L)	-4.67 (-12.84 , 3.5) p=0.228	-1.71 (-7.4 , 3.98) p=0.514	-0.67 (-6.13 , 4.8) p=0.789	1.96 (-0.8 , 4.72) p=0.142	0.45 (-5.32 , 6.22) p=0.865	-0.63 (-4.72 , 3.45) p=0.734
Posterior thalamic radiation (R)	-3.04 (-13.69 , 7.61) p=0.535	-1.73 (-8.75 , 5.29) p=0.591	1.48 (-5.15 , 8.1) p=0.626	1.7 (-1.92 , 5.31) p=0.316	1.35 (-5.66 , 8.35) p=0.674	-0.73 (-5.73 , 4.28) p=0.751
Posterior thalamic radiation (L)	-4.51 (-13.87 , 4.85) p=0.304	-1.75 (-8.14 , 4.64) p=0.551	2.62 (-3.2 , 8.44) p=0.335	-0.43 (-3.92 , 3.06) p=0.789	-0.17 (-6.64 , 6.29) p=0.953	-1.29 (-5.79 , 3.21) p=0.532
Sagittal stratum (R)	-7.37 (-17.6 , 2.85) p=0.137	-2.59 (-9.96 , 4.79) p=0.448	-0.28 (-7.46 , 6.89) p=0.931	-0.01 (-4.11 , 4.09) p=0.995	-1.18 (-8.69 , 6.33) p=0.731	-1.93 (-7.11 , 3.25) p=0.421
Sagittal stratum (L)	-7.13 (-19.18 , 4.93) p=0.214	-2.57 (-11 , 5.87) p=0.508	-1.22 (-9.31 , 6.87) p=0.741	-2.07 (-6.45 , 2.3) p=0.312	-4.94 (-12.66 , 2.78) p=0.182	-1.32 (-7.34 , 4.7) p=0.631
External capsule (R)	-9.35 (-26.43 , 7.74) p=0.247	-9.92 (-19.44 , -0.39) p=0.043	-5.9 (-16.4 , 4.59) p=0.235	4.58 (-0.93 , 10.09) p=0.093	-3.11 (-14.88 , 8.66) p=0.564	-3.29 (-11.46 , 4.88) p=0.386
External capsule (L)	-11.38 (-25.06 , 2.3) p=0.093	-9.09 (-17.16 , -1.03) p=0.031	-5.5 (-14.55 , 3.57) p=0.203	3.4 (-1.68 , 8.48) p=0.164	-2.71 (-13 , 7.59) p=0.567	-3.7 (-10.62 , 3.22) p=0.257
Cingulum cingulate (R)	-1.93 (-9.63 , 5.77) p=0.585	-4.47 (-8.35 , -0.59) p=0.029	-2.12 (-6.68 , 2.44) p=0.321	2.44 (0.38 , 4.5) p=0.025	0.48 (-4.6 , 5.56) p=0.836	-1.3 (-4.79 , 2.19) p=0.422
Cingulum cingulate (L)	-7.42 (-18.14 , 3.3) p=0.152	-9.03 (-13.08 , -4.98) p=0.001	-6.04 (-11.94 , -0.14) p=0.046	2.83 (-0.85 , 6.51) p=0.116	-0.84 (-8.66 , 6.99) p=0.815	-3.8 (-8.59 , 1) p=0.107
Cingulum hippocampus (R)	0.9 (-9.74 , 11.55) p=0.852	-3.18 (-9.75 , 3.39) p=0.302	0.17 (-6.4 , 6.75) p=0.954	1.2 (-2.44 , 4.85) p=0.474	-3.05 (-9.59 , 3.48) p=0.318	1.1 (-3.76 , 5.96) p=0.621
Cingulum hippocampus (L)	-3.43 (-8.28 , 1.41) p=0.144	-3.00 (-5.81 , -0.2) p=0.038	-0.8 (-4.14 , 2.53) p=0.599	0.47 (-1.43 , 2.37) p=0.589	-0.18 (-3.74 , 3.39) p=0.914	-1.68 (-3.88 , 0.52) p=0.118
Fornix cres Stria terminalis (R)	-1.52 (-9.46 , 6.42) p=0.675	-2.44 (-7.37 , 2.48) p=0.291	-2.21 (-6.87 , 2.45) p=0.312	0.92 (-1.82 , 3.66) p=0.467	0.01 (-5.2 , 5.23) p=0.996	-0.93 (-4.57 , 2.71) p=0.576
Fornix cres Stria terminalis (L)	-2.51 (-9.53 , 4.5) p=0.439	-3.23 (-7.32 , 0.87) p=0.109	-3.54 (-7.14 , 0.06) p=0.053	0.98 (-1.47 , 3.43) p=0.39	-0.71 (-5.4 , 3.98) p=0.74	-1.47 (-4.64 , 1.7) p=0.322
Superior longitudinal fasciculus (R)	-6.29 (-15.22 , 2.63) p=0.145	-4.46 (-10.17 , 1.24) p=0.111	-4.31 (-9.63 , 1.01) p=0.1	1.82 (-1.47 , 5.1) p=0.242	-4.03 (-9.85 , 1.8) p=0.152	-1.44 (-5.98 , 3.1) p=0.492
Superior longitudinal fasciculus (L)	-1.66 (-12.25 , 8.92) p=0.731	-4.92 (-10.84 , 1) p=0.093	-2.44 (-8.75 , 3.87) p=0.404	2.56 (-0.67 , 5.77) p=0.106	-1.38 (-8.23 , 5.46) p=0.659	-0.76 (-5.65 , 4.13) p=0.733
SFOF (R)	-1.51 (-9.34 , 6.33) p=0.674	-0.19 (-5.38 , 4.99) p=0.934	-1.42 (-6.19 , 3.34) p=0.516	-2.11 (-4.4 , 0.18) p=0.067	-2.85 (-7.52 , 1.83) p=0.202	-0.49 (-4.13 , 3.15) p=0.769
SFOF (L)	-9.59 (-18.84 , -0.33) p=0.044	-5.86 (-12.16 , 0.44) p=0.065	-5.7 (-11.53 , 0.13) p=0.054	-0.58 (-4.69 , 3.53) p=0.758	-6.21 (-12.23 , -0.18) p=0.045	-2.28 (-7.43 , 2.86) p=0.342
Uncinate fasciculus (R)	0.00 (-5.73 , 5.73) p=0.999	-0.01 (-3.77 , 3.74) p=0.994	-0.37 (-3.89 , 3.15) p=0.818	-0.37 (-2.37 , 1.63) p=0.685	-0.29 (-4 , 3.43) p=0.866	-0.55 (-3.16 , 2.06) p=0.646
Uncinate fasciculus (L)	1.81 (-5.46 , 9.08) p=0.586	-0.65 (-5.47 , 4.18) p=0.769	-0.77 (-5.3 , 3.75) p=0.708	1.42 (-0.95 , 3.8) p=0.208	-1.56 (-6.22 , 3.1) p=0.469	1.17 (-2.14 , 4.47) p=0.444
Estimates represent expected change in number of errors within each emotion per 0.1 unit increase in FA. Dark grey highlights significant associations (p<0.05), light grey highlights associations of borderline significance (p<0.08).						

The HD group performance levels in disgust, anger and fear recognition were associated with notably more regions of decreased FA than for the other emotions (illustrated in Figure 14-1). Poorer disgust recognition was associated with reduced FA within the cerebral peduncle (right significantly and left of borderline significance), anterior corona radiata, external capsule, cingulum, left SFOF (borderline significance) and CC. Poorer anger recognition was associated with reduced FA within the left cerebral peduncle, posterior internal capsule and left SFOF. Impaired fear recognition was associated with reduced FA within the anterior internal capsule, right anterior corona radiata and left cingulum. The other three emotions (which were also found to be less impaired compared with the control group performance levels) showed fewer significant FA associations. Significant associations were however found between poorer happiness recognition with increased right cingulum FA; sadness recognition impairments with decreased left anterior internal capsule and left SFOF FA; and surprise with left cerebral peduncle FA.

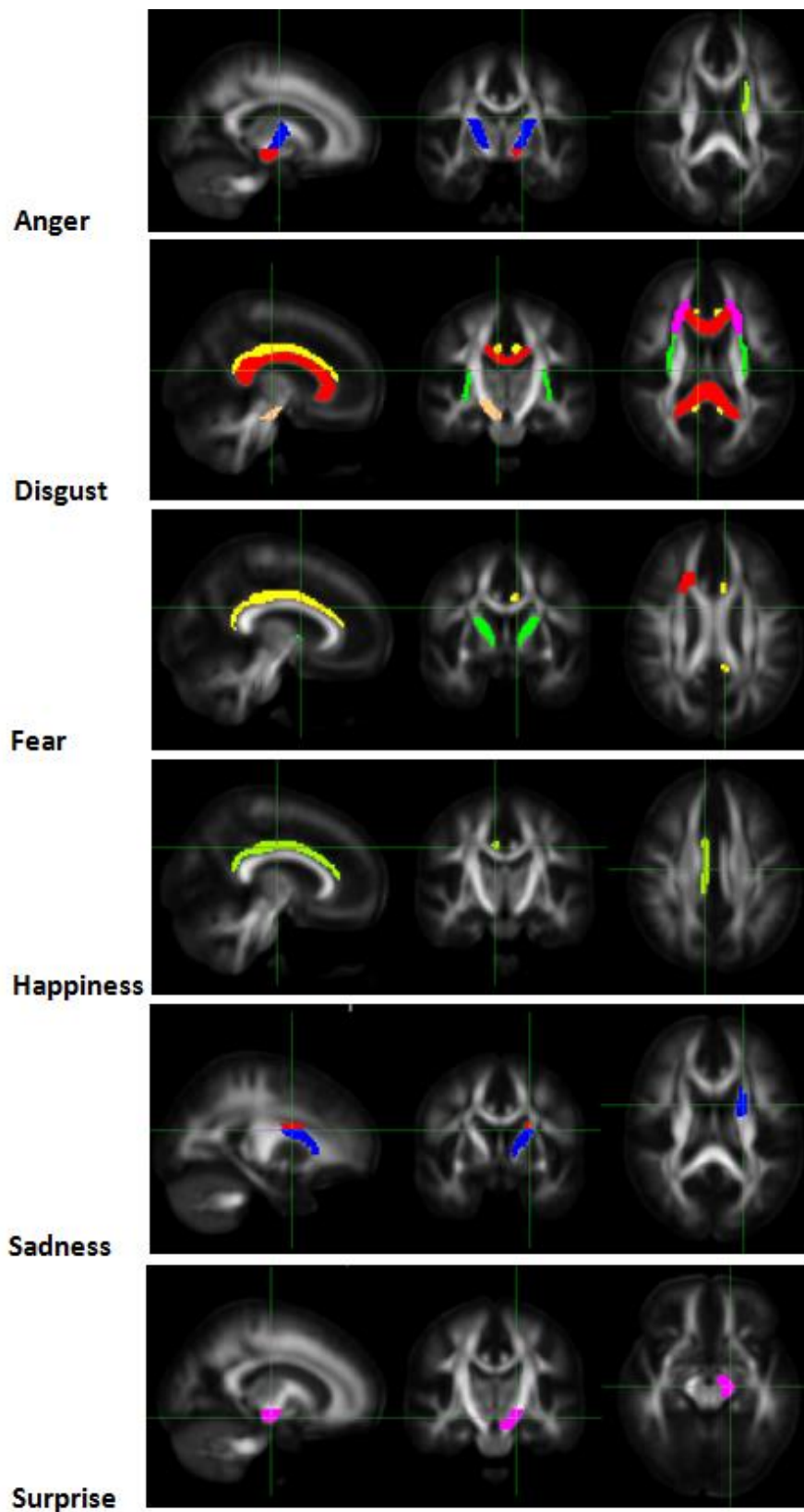


Figure 14-1. WM tracts in which FA significantly associated with recognition of anger, disgust, fear, happiness, sadness and surprise (top to bottom) in the HD group. Colours are random for each region and do not represent any strength to the observed associations.

14.4 Discussion

There is currently no consensus as to whether the disproportionate emotion-specific recognition deficits in HD are a result of damage to distinct and separable neural substrates (Adolphs 2002; Sprengelmeyer et al. 1998) or a more general, interlinked emotion network in which specific emotions share some, but not all, neural substrates (Henley et al. 2008). In a novel task, recognition of disgust, anger and fear was found to be disproportionately impaired in early-stage HD, although to differing extents in different stimulus modalities (photos, vocal expressions and film clips). This exploratory imaging investigation into the underlying neural correlates of this impairment found significant associations between task performance and reduced volume, thickness and FA within widespread GM regions and WM tracts. The patterns of these associations did not however fulfil expectations from previous literature of emotion-specific separable neural substrates (Adolphs 2002). There also did not seem to be a consistent pattern of macro- or micro-structural pathology underlying emotion recognition via different stimulus modalities. Overall these findings are indicative of a widespread emotion network, differentially recruited by varying task demands and suffering from degradation of connective WM integrity, and to a lesser extent regional GM volume loss, in early-stage HD.

Of the subcortical regions, only caudate volume showed any real significant association with task performance. Due to the disproportionate atrophy in this region in HD it is arguable that this may reflect general disease progression and cognitive decline rather than specific emotion recognition deficits. Thickness of the cortex within the fusiform and superior temporal gyri associated significantly with the number of errors in fear recognition and the insula was significantly associated with the number of sadness recognition errors. Overall these results do not fulfil expectations from previous literature for associations between: the insula and basal ganglia and disgust processing (Adolphs 2002; Sprengelmeyer et al. 1998), the cingulate gyrus, left medial temporal gyrus, ventral striatum and anger, the amygdala and fear (Adolphs 2002), the fusiform gyrus and facial processing (Kanwisher et al. 1997), and the STG and perception of expressive movements (Puce et al. 1998). This may be due to the relatively small sample size, methodology used or the strength of functional compared with morphometric relationships (these associations were derived from functional imaging studies).

There was no detectable common regional abnormality underlying the three more impaired emotions (disgust, anger and fear) but performance in all three was significantly associated with more regions of reduced WM FA than the three unimpaired emotions. FA is a measure of the coherence of water diffusion through the fibrous WM. Decreased FA suggests disorganisation and/or reduced structural integrity of WM tracts. This analysis highlighted six WM regions that appeared to be related to recognition of several emotions and therefore may constitute the WM 'core' of the hypothesized emotion network. These were: the cerebral peduncles, internal capsule; anterior corona radiata, SFOF, cingulum and CC.

The **cerebral peduncles** are found in the anterior midbrain at the base of the **internal capsule**. The internal capsule then branches upwards and outwards to form the **corona radiata**. The significant associations between emotion recognition performance and FA within these three WM tracts therefore highlights one large branching structure which is associated, in certain regions, with all the tested emotions except happiness. As the name suggests, the **SFOF** runs from the frontal to the occipital cortex. Tract-specific analysis of the SFOF in schizophrenia found this tract to connect the prefrontal area to the thalamus and its FA to be deteriorated compared with healthy controls (Kunimatsu et al. 2008). In the current study significant results were found between reduced SFOF FA and impaired anger and sadness recognition with borderline associations with disgust, fear and happiness. The **cingulum** provides substantial WM connections within the cortico-limbic neural system that subserves emotional regulation and expression (Devinsky et al. 1995). Abnormalities of the structural integrity of the anterior cingulum have been linked to depression (Cole et al. 2012), apathy (Kim et al. 2011) and bipolar disorder (Wang et al. 2008). Poorer recognition of disgust and fear was found to be associated with decreased FA within the cingulum whereas increased FA was associated with poorer happiness recognition. This particular association of increased FA and poorer performance is unexpected and on closer examination seems to be driven by the adjustment for motor response time. This did not seem to be the case for other associations however, and since motor response is considered a potential confounder it does seem appropriate to include it in this model. The **CC** is the communication bridge between the two hemispheres of the brain. Decreased FA within all three regions of the CC (genu, body and splenium) was significantly associated with disgust recognition which suggests an inter-hemispheric emotion (disgust) network.

The fMRI literature does reveal a commonality between the three impaired emotions (disgust, anger and fear). These emotions have consistently been associated with brain activations within distinct brain regions (Adolphs 2002); surprise, sadness and happiness perception have not and therefore may require less specific network activation. It is hypothesized that the seemingly higher reliance on the WM micro-structure and specific neuronal activation patterns may make disgust, anger and fear more vulnerable to brain changes within this emotion network. Additionally, the widespread nature of the network could make it particularly susceptible to disruptions in WM connectivity. This hypothesis is however speculative until reproduced by larger studies.

This study was limited by sample size and these exploratory imaging findings would clearly benefit from replication. It should also be noted that multiple statistical tests were conducted with no adjustment for multiple comparisons. Due to the clinical population available for this study there was an imbalance in males and females within the HD group. This was accounted for in all statistical analyses by adjusting for gender. Volumetric analyses were also adjusted for TIV, known to vary between genders. The image analysis involved removal of group template voxels with FA>0.2 from the regions. It should be noted that

this could have introduced an element of circularity between method and outcome, i.e. the method is altered according to the results. Interpretation of the imaging associations is limited because without jointly modelling the between-group difference emotion recognition data with the imaging associations it is not possible to conclude that the observed deficit in ability compared with controls is a consequence of the associations with atrophy or diffusion abnormalities in the HD patients. Nevertheless the exploration of associations between emotion recognition performance and micro- as well as macro-structural brain measures in HD makes this the first study to combine these analyses.

Deficits in disgust, anger and fear recognition were found in this relatively early-stage HD cohort; potentially indicative of a common impairment or damaged process among these three emotions. This exploratory imaging analysis did not however detect any shared pattern of macro- or micro-structural brain abnormalities, but performance on the affected emotions did associate with reduced FA within more extensive WM regions than the unaffected emotions. These results are suggestive of a substantial influence of WM connectivity on emotion processing in HD. The detection of associations across the whole-brain implicates a widespread emotion network, potentially employed in differential ways depending on the task demands. Further work is needed to build on this exploratory imaging investigation, perhaps by combining functional and diffusion analyses. It is however clear that a whole-brain approach is required to fully characterise the emotion network.

15 A Silent Contribution of the Visual Cortex to Cognitive Task

Performance: An Investigation in HD

15.1 Introduction

The occipital cortex is primarily viewed as the visual area of the brain, with topologically distinct regions believed to manage different aspects of visual processing (Wandell et al. 2007). In HD atrophy occurs in the occipital lobe in the absence of any apparent impairment in visual function (Tabrizi et al. 2009). Cognitive deficits are however detectable in HD on multiple tasks designed to probe executive functioning (Cabeza and Nyberg 2000). Executive function was traditionally attributed to frontal lobe activations (Reitan and Wolfson 1994). More recent evidence however suggests that this is an over-simplification, with frontal activations necessary but not fully sufficient for explaining task performance (Alvarez and Emory 2006). Since many cognitive tests involve a visual component, for example written responses or symbol identification, impaired visual processing due to occipital cortex atrophy could affect performance on these tests. This study aimed to investigate whether, and to what extent, thickness of the visual cortex associates with cognitive test performance. This association was studied in HD in which there are typically no overt visual deficits but occipital atrophy is present.

Specific regions of the occipital cortex are known to be associated with different visual processing functions; some basic and others more complex. Therefore the location of occipital cortex pathology is likely to affect the manifestation of any visual deficit and consequently its potential impact on cognitive test performance. Current understanding of the functional topology of the occipital lobe, which is derived from visual field mapping and supported by functional imaging studies (see Cabeza and Nyberg (2000) for a review), divides the cortex into a number of processing areas that extend over the occipital gyri. The primary visual area (V1), also called Brodmann Area (BA) 17, is located within the pericalcarine region and maps the visual information in view at any one time; known as the visual field (Tootell et al. 1998; Wandell et al. 2007). This visual information is then projected to higher-level processing regions (Roe and Ts'o 1995), known as V2, V3 and V4. Visual area 2 (V2, BA18) is located within the cuneus above, and the lingual gyrus below V1 and is thought to process colour as well as computing orientation and disparity information (Roe & Ts'o 1995; Zeki 1978). Visual area 3 (V3), believed to play a role in motion perception (Gegenfurtner et al. 1997; Larsson and Heeger 2006), is also located within the cuneus but extends laterally into the lateral occipital cortex (LOC). The major functional contribution of the LOC is thought to be integration of visual information, especially shape information, as well as performing visual and tactile object recognition (Beauchamp 2005; Grill-Spector et al. 2001; Larsson & Heeger 2006). Within the lingual gyrus is visual area 4 (V4), believed to be a colour processing area (Tootell et al. 2003). The lingual region has also been reported to be involved in processing visual memory and identifying words and letters (Mechelli et al. 2000). To

summarize, visual areas V1 to V4, described here in relation to four regions of the occipital cortex (the cuneus, pericalcarine, lingual and LOC) categorise the topologically-specific functionality in the occipital lobe. It is currently unclear what effect pathology within these regions has on performance levels in cognitive tasks with visual components. This will be investigated in HD.

Whilst visual deficits are not a known feature of the HD phenotype, several imaging studies have detected occipital cortex atrophy in both pre- and manifest HD (Rosas et al. 2008;Stoffers et al. 2010;Tabrizi et al. 2009;Wolf et al. 2014) and there is a marked reduction in neural number in the occipital lobe at post-mortem (Lange et al. 1976). An HD cohort ranging from pre- to manifest disease would therefore be expected to exhibit both varying levels of cognitive impairment and occipital atrophy, with no notable visual deficit. These characteristics mean that HD provides an opportunity to investigate the contribution of occipital cortex pathology to performance on a range of cognitive tasks involving a visual component.

This study aimed to test associations between thickness of the visual cortex and cognitive test performance. Cortical thickness in four occipital regions (the cuneus, pericalcarine, lingual and LOC) was compared between healthy controls, pre-manifest and manifest HD participants, and the relationship between occipital lobe thickness and cognitive impairment was examined in the HD gene-carriers. It was hypothesized that the thickness of the lingual and LOC, which are known to be involved in higher-level visual functioning such as word reading, visual memory, integration of visual information and object recognition, would associate with cognitive test performance. In contrast, as vision is not affected in HD, it was hypothesized that the regions responsible for more basic visual processing, notably the pericalcarine gyrus and cuneus, would not show significant associations with cognitive test performance.

15.2 Methods

15.2.1 Cohort

Participants were recruited across four study sites as part of the TRACK-HD study (Tabrizi et al. 2009), 313 of which were included in the present investigation of the occipital cortex (107 controls, 116 preHD and 90 HD individuals). The preHD cohort was separated into two groups; those estimated to be more than 10.8 years from disease onset (Langbehn et al. 2010) were classified as the preHD-A group (n=54) and those less than 10.8 years from estimated onset, preHD-B (n=62). Using the UHDRS (1996) the HD cohort was classified based on their TFC scores as HD1 (n=51, TFC=11-13) or HD2 (n=39, TFC=7-10). The control group comprised of partners, spouses or gene-negative siblings of the gene-carriers. Full selection criteria and data collection processes have been previously published (Tabrizi et al. 2009;Tabrizi et al. 2012).

Participants were tested for visual function using a Snellen Visual Acuity equivalent, the Low-Contrast SLOAN Letter Charts (Balcer et al. 2000). This provided a score ranging from 1-12, with 1 representing poor

visual acuity (20/200 vision) and 12 representing high acuity (20/16 vision). Participants with a score of less than 11, which is equivalent to below the average 20/20 vision, were excluded from the current study. This resulted in a final cohort of 275 participants (97 healthy controls, 51 preHD-A, 58 preHD-B, 40 HD1 and 29 HD2 participants).

15.2.2 Clinical Assessments

A selection of widely used cognitive tests with visual components was selected from the larger TRACK-HD study battery (Tabrizi et al. 2009). One motor task without a visual component was also included. This was the Paced Tapping task (Reilmann et al. 2005) which has been found to be a significant predictor of disease progression in HD (Tabrizi et al. 2013) and so was included as a general proxy for disease progression (with no visual component) to test the exclusivity of the visual-occipital relationship.

1. *Symbol Digit Modalities Test, SDMT* (Smith 1991). The SDMT is a pencil-and-paper task that assesses visuomotor integration and has components of visual scanning and tracking. The participant has 90 seconds to match as many symbols with the corresponding digits as possible, with the final score being the number of correctly matched symbols.
2. *Stroop Word Reading* (MacLeod 1991). This Stroop task requires visual scanning, cognitive control and processing speed. Participants must read as many names of colours from a presented card as possible in 45 seconds and are scored by the number of words correctly read out loud.
3. *Trails A* (Sprensen and Strauss 1985). The Trail Making Task measures sustained attention and information processing speed. Participants are timed whilst drawing an ordinal line through randomly dispersed numbers from 1 to 25. The final score is the time taken to complete the task.
4. *Map Search* (Robertson et al. 1994). This is a test of visuospatial selective attention. Participants search a visually cluttered display to identify target symbols amongst distractors and are scored by the total number of symbols identified in two minutes.
5. *Mental Rotation* (Shepard and Metzler 1971). The mental rotation task assesses spatial processing. Participants are asked to compare two figures and determine whether they are identical but rotated, or mirror image figures. The total score is the percentage of items answered correctly.
6. *Spot the Change* (Cowan et al. 2005). This test examines the participant's ability to sustain object and location representations without rehearsal. An array of coloured squares (either 4 or 8) is briefly presented for 250ms. This is followed by an identical

array, where one of the squares is encircled. The participant has to decide whether that square has changed colour. The total score is the percentage answered correctly.

7. *Paced Tapping* (Reilmann et al. 2005). The Paced Tapping task is a measure of motor performance and control. Participants are required to tap at the same speed as a previously played 1.8 Hz tone and deviation in tapping pace compared to this tone is assessed.

15.2.3 Image Acquisition

3T T1-weighted scans were acquired from four scanners utilising previously validated protocols for multi-site use (Tabrizi et al. 2009); details in Appendix Section 23.3. All images were visually assessed for quality; specifically artefacts such as motion, distortion and poor tissue contrast (IXICO Ltd. and TRACK-HD imaging team, London, UK). T1-weighted scans were then bias-corrected using the non-parametric non-uniform intensity normalization (N3) method of Sled et al. (Sled et al. 1998), with optimised parameters for 3T data as outlined in Boyes et al. (Boyes et al. 2008).

15.2.4 Image Analysis

The analysis was run using the default version 5.3.0 FreeSurfer pipeline with the -3T flag to optimise analysis for 3T data (Fischl et al. 2002; Fischl et al. 2004; Fischl & Dale 2000). Outputs were visually inspected. On a small number of scans there was slight overestimation of the occipital GM-CSF boundary, however this was deemed to be minor and therefore no participants were excluded from the analysis.

The cortical thickness analysis was limited to the four occipital regions defined by the Desikan-Killiany atlas (Desikan et al. 2006): the cuneus, pericalcarine, lingual and LOC (shown in Figure 15-2A). Two separate image analyses were conducted. Firstly, average thickness values for the four regions were extracted from each scan. An averaged occipital lobe thickness estimate was also calculated using a scaling factor to account for differing surface areas of the four regions; mean thickness for each region was multiplied by the surface area of that region and the sum of these four regions was divided by the total surface area. This calculation was also used to account for surface area when averaging the regional results across hemispheres. Secondly, statistical maps of the clinical associations were also processed in FreeSurfer. For this, data from all HD gene-carriers was pooled to create a study-specific group template. Average thickness data was overlaid on this template and smoothed with a 10mm full width at half maximum kernel before clinical associations were computed.

An averaged frontal cortex thickness estimate was also extracted from the data. This was included as a covariate in regression models fitted between cortical thickness and cognitive task performance data in order to isolate the effect of occipital cortex atrophy above that of frontal (executive processing) pathology.

15.2.5 Statistical Analysis

Generalised least squares regression models (adjusted for age, gender and study site) were fitted to the regional thickness data to assess between-group differences. Relationships between cognitive test performance and regional cortical thickness in the HD gene-carriers were assessed using general linear regression models, adjusting for age, gender, study site, education, CAG, disease burden score and frontal cortex thickness. This was repeated five times for each cognitive test to examine the relationships between performance levels and each of the five occipital regions (average occipital cortex, pericalcarine, cuneus, lingual and LOC). R-squared (R^2) was used to determine the contribution of occipital thickness to each regression model, i.e. the proportion of variance in each task accounted for by occipital thickness. This was calculated using this formula:

$$1 - [(residual\ sum\ of\ squares)/(total\ sum\ of\ squares)]$$

A FreeSurfer-programmed 'Different Onset, Same Slope' (DOSS) general linear model was fitted to the spatially normalised thickness data to create statistical maps displaying the relationships between cortical thickness and cognitive test performances in the HD gene-carriers. These models were adjusted for age, gender, study site, education, CAG, disease burden score and frontal cortex thickness. The results were corrected for multiple comparisons using a Monte Carlo cluster-wise simulation with a vertex-wise threshold of $p < 0.001$ (two-tailed) and a cluster-wise threshold of $p < 0.025$ for each hemisphere, limited to the occipital region.

15.3 Results

Table 15-1. Demographic data. Data are presented as mean (SD) range.							
	Controls (n=97)	PreHD-A (n=51)	PreHD-B (n=58)	Combined (n=109)	HD1 (n=40)	HD2 (n=29)	Combined (n=69)
Age	47.8 (9.8) 25.1-67.6	41.4 (9.77) 20.5-61.4	43.3 (8.4) 28.4-66.1	42.4 (9.1) 20.5-66.1	48.4 (10.3) 31.0-65.7	49.3 (8.7) 34.4-64.4	48.8 (9.6) 31.0-65.7
Women	53.6%	58.8%	53.5%	56.0%	67.5%	51.7%	60.9%
Education (ISCED^a)	4.0 (1.3) 1-6	4.2 (1.1) 2-6	4.0 (1.1) 2-6	4.1 (1.1) 2-6	3.8 (1.4) 2-6	3.2 (1.0) 2-5	3.6 (1.2) 2-6
Disease Burden Score^b	--	260.2 (29.4) 177.6-304.3	339.9 (38.6) 282.5-489.5	302.6 (52.8) 177.6-489.5	367.3 (76.9) 217.4-574.1	392.6 (87.6) 162.8-583.0	378.0 (81.9) 162.8-583.0
CAG	--	42.2 (2.1) 39-49	43.6 (2.1) 40-50	43.0 (2.2) 39-50	43.6 (2.8) 39-50	43.8 (2.7) 39-52	43.7 (2.8) 39-52
Visual Acuity^c	11.9 (0.4) 11-12	11.94 (0.2) 11-12	11.8 (0.4) 11-12	11.9 (0.4) 11-12	11.8 (0.4) 11-12	11.7 (0.5) 11-12	11.7 (0.5) 11-12
^a ISCED = International Standard Classification of Education; ^b Disease burden formula (Penney et al. 1997) = (CAG-35.5) × age; ^c Snellen visual acuity scale: 1 to 12 represents very poor to high visual acuity respectively.							

Participant demographics are detailed in Table 15-1. The three groups were well matched, although the preHD group was, as expected, on average significantly younger than the other groups ($p < 0.05$). A steady decrease in cortical thickness with increasing disease burden was seen in all occipital regions across the four gene-carrier groups compared to controls (Figure 15-1).

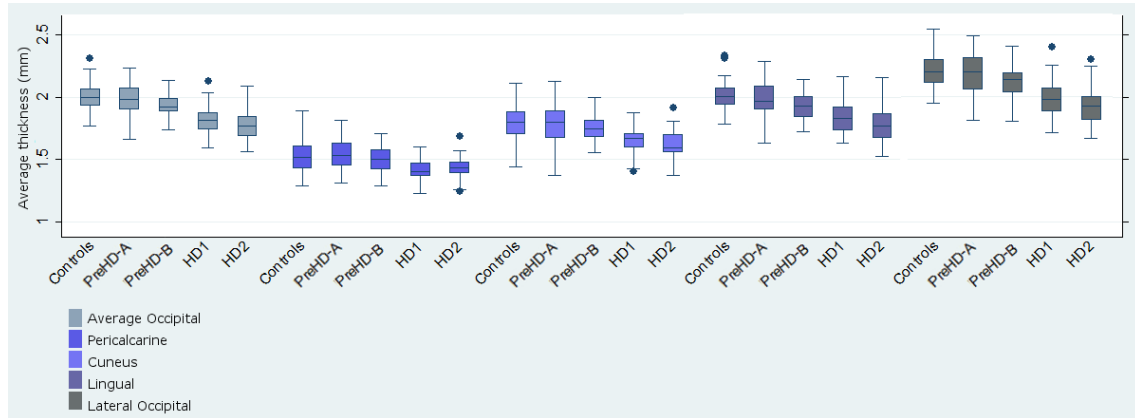


Figure 15-1. Occipital cortex thickness (mm) plotted by subregion and subgroup. The box plot whiskers range from the minimum to the maximum values (excluding outliers), the box spans from the 25th to the 75th percentile and the central line represents the median value.

This reduced thickness in HD compared with healthy controls is significant ($p < 0.05$) as early as the preHD-A group (> 10.8 years from estimated onset) in the averaged occipital cortex, the lingual gyrus and the LOC; details of between-group differences are provided in Table 15-2. In preHD individuals estimated to be < 10.8 years from onset (preHD-B), as well as the HD1 and HD2 individuals, the cortical thickness was significantly lower than controls in all four occipital regions as well as the averaged occipital cortex. All regions showed significantly thinner cortices in the HD1 group compared with preHD-B participants. No significant thickness differences were found between HD1 and HD2 groups.

Table 15-2. Adjusted between-group differences in occipital cortical thickness measures; reported with p-values and effect sizes.

Thickness (mm)							
	PreHD vs Controls	HD vs Controls	PreHD vs HD	PreHD-A vs Controls	PreHD-B vs Controls	HD1 vs PreHD-B	HD2 vs HD1
Occipital cortex	-0.067 (-.093, -.040) p<.001 ES = -.662 (-.936, -.358)	-0.107 (-.122, -.093) p<.001 ES = -1.361 (-1.757, -.960)	-0.107 (-.122, -.093) p<.001 ES = -.963 (-1.169, -.748)	-0.046 (-.083, -.009) p=.015 ES = -.400 (-.738, -.056)	-0.041 (-.055, -.027) p<.001 ES = -.511 (-.709, -.308)	-.111 (-.148, -.075) p<.001 ES = -1.094 (-1.520, -.614)	-.055 (-.102, -.008) p=.022 ES = -.501 (-1.047, .043)
LOC	-.077 (-.112, -.044) p<.001 ES = -.589 (-.866, -.308)	-.139 (-.158, -.120) p<.001 ES = -1.363 (-1.771, -.939)	-.202 (-.244, -.160) p<.001 ES = -.939 (-1.152, -.719)	-.055 (-.102, -.008) p=.023 ES = -.374 (-.715, -.020)	-.047 (-.066, -.029) p<.001 ES = -.423 (-.608, -.226)	-.154 (-.203, -.104) p<.001 ES = -1.141 (-1.619, -.626)	-.070 (-.132, -.007) p=.028 ES = -.483 (-1.006, .082)
Lingual	-.073 (-.102, -.044) p<.001 ES = -.663 (-.927, -.386)	-.101 (-.118, -.084) p<.001 ES = -.942 (- 1.282, -.591)	-.125 (-.162, -.090) p<.001 ES = -.758 (-1.049, -.467)	-.053 (-.090, -.016) p=.001 ES = -.476 (-.822, -.122)	-.045 (-.061, -.028) p<.001 ES = -.434 (-.602, -.253)	-.071 (-.116, -.026) p=.002 ES = -.594 (-1.026, -.169)	-.072 (-.129, -.016) p=.012 ES = -.540 (-1.049, -.000)
Cuneus	-.052 (-.085, -.019) p=.002 ES = -.441 (-.722, -.122)	-.078 (-.094, -.062) p<.001 ES = -1.113 (-1.508, -.751)	-.120 (-.153, -.086) p<.001 ES = -.709 (-1.084, -.334)	-.033 (-.079, -.013) p=.161 ES = -.237 (-.585, .128)	-.033 (-.050, -.017) p<.001 ES = -.379 (-.564, -.177)	-.086 (-.127, -.046) p<.001 ES = -.781 (-1.168, -.397)	-.033 (-.078, .011) p=.143 ES = -.353 (-.942, .208)
Pericalcarine	-.028 (-.060, .004) p=.085 ES = -.272 (-.568, .058)	-.051 (-.066, -.036) p<.001 ES = -.944 (-1.319, -.594)	-.088 (-.117, -.058) p<.001 ES = -.560 (-.769, -.357)	-.017 (-.058, .024) p=.423 ES = -.148 (-.522, .244)	-.019 (-.034, -.002) p=.033 ES = -.204 (-.397, -.012)	-.075 (-.111, -.039) p<.001 ES = -.863 (-1.339, -.357)	-.002 (-.041, .037) p=.919 ES = -.022 (-.529, .450)
Data are coefficients (95% CI), p-values and effect sizes (bootstrapped 95% CIs) of between-group differences in occipital cortical thickness measures. Coefficients and p-values calculated using generalised least squares regression adjusted for age, gender and study site. Effect sizes (ES) were calculated as the estimated absolute adjusted mean difference of the metric between the HD and control groups, divided by the estimated residual SD of the HD group. These are reported with bias-corrected and accelerated bootstrapped 95% CIs based on 2000 replications (Carpenter & Bithell 2000).							

All six cognitive tasks showed statistically significant associations ($p < 0.05$) with averaged occipital cortex thickness in the expected direction, e.g. reduced thickness associated with poorer performance (Table 15-3). The SDMT showed the highest association, with average occipital cortex thickness accounting for an estimated 7.5% of the variance in task performance. When the four regions of the occipital cortex were examined separately, the LOC and lingual regions showed the greatest associations with task performance. The LOC thickness explained an estimated 9.5% of the variance in SDMT performance, followed by 6.7% for both the Stroop Word Reading and Spot the Change tasks. Thickness of the lingual region showed smaller associations with cognitive task performance; the greatest association found with performance in Spot the Change (~6.1% of variance explained). The only significant association between cognitive test performance and cuneus thickness was for the Stroop Word Reading test, with just ~1.8% of variance explained. None of the tasks were associated with pericalcarine thickness. There were also no significant associations between occipital cortex thickness and Paced Tapping performance, suggesting that these significant associations are specific to visual cognition rather than general disease progression.

Table 15-3. Associations, in the HD group, between cortical thickness within occipital regions and cognitive and motor task performance. Light grey highlights statistically significant ($p < 0.05$) associations.

	Average Occipital Thickness	Pericalcarine	Cuneus	Lingual	LOC
Paced Tapping (n=178)	-50.500 (-153.29, 52.30) $p = .334$ $R^2 = .005$	-24.013 (-127.21, 79.18) $p = .647$ $R^2 = .001$	-60.466 (-145.89, 24.96) $p = .164$ $R^2 = .010$	12.860 (-72.88, 98.60) $p = .768$ $R^2 = .000$	-53.979 (-132.47, 24.51) $p = .176$ $R^2 = .010$
SDMT (n=173)	42.537 (26.08, 59.00) $p < .001$ $R^2 = .075$	14.712 (-3.35, 32.77) $p = .110$ $R^2 = .009$	10.045 (-4.80, 24.89) $p = .183$ $R^2 = .006$	28.813 (14.67, 42.96) $p < .001$ $R^2 = .049$	36.502 (24.19, 48.81) $p < .001$ $R^2 = .095$
Stroop Word Reading (n=173)	58.571 (29.66, 87.49) $p < .001$ $R^2 = .056$	21.25 (-9.69, 52.18) $p = .177$ $R^2 = .007$	28.013 (2.87, 53.16) $p = .029$ $R^2 = .018$	32.217 (7.36, 57.07) $p = .011$ $R^2 = .024$	49.080 (27.21, 70.95) $p < .001$ $R^2 = .067$
Trails A (n=173)	-35.195 (-54.69, -15.70) $p < .001$ $R^2 = .051$	-15.738 (-36.38, 4.90) $p = .134$ $R^2 = .010$	-10.955 (-27.92, 6.01) $p = .204$ $R^2 = .007$	-31.732 (-47.93, -15.53) $p < .001$ $R^2 = .059$	-25.992 (-40.92, -11.06) $p = .001$ $R^2 = .047$
Map search (n=170)	33.742 (12.53, 54.95) $p = .002$ $R^2 = .034$	1.034 (-21.51, 23.58) $p = .928$ $R^2 = .000$	8.183 (-10.34, 26.71) $p = .384$ $R^2 = .003$	22.509 (4.59, 40.43) $p = .014$ $R^2 = .022$	31.053 (15.13, 46.98) $p < .001$ $R^2 = .050$
Mental Rotation (n=173)	.216 (.00, .43) $p = .045$ $R^2 = .017$.104 (-.11, .32) $p = .349$ $R^2 = .004$.062 (-.16, .20) $p = .790$ $R^2 = .000$.075 (-.10, .25) $p = .410$ $R^2 = .003$.213 (.05, .37) $p = .009$ $R^2 = .028$
Spot the Change (n=178)	36.457 (18.33, 54.59) $p < .001$ $R^2 = .064$	12.566 (-6.33, 31.47) $p = .191$ $R^2 = .008$	5.777 (-10.01, 21.56) $p = .471$ $R^2 = .002$	29.539 (14.42, 44.67) $p < .001$ $R^2 = .061$	28.542 (14.69, 42.39) $p < .001$ $R^2 = .067$
Associations assessed using general linear regression, adjusted for gender, age, study site, education, CAG, disease burden score and frontal lobe thickness. R^2 explains the proportion of variance accounted for by each region and was calculated by $1 - (\text{residual sum of squares} / \text{total sum of squares})$.					

The significance maps shown in Figure 15-2B provide further information about the location of these structure-function associations. For all tasks the significant clusters within the LOC appeared to be mostly in the occipital pole. For SDMT, Stroop Word Reading, and Trails A the lingual cluster was lateralised to the right hemisphere, but for Spot the Change there were bilateral associations within the lingual region. Consistent with the associations shown in Table 15-3, there were no significant clusters within the pericalcarine or cuneus regions.

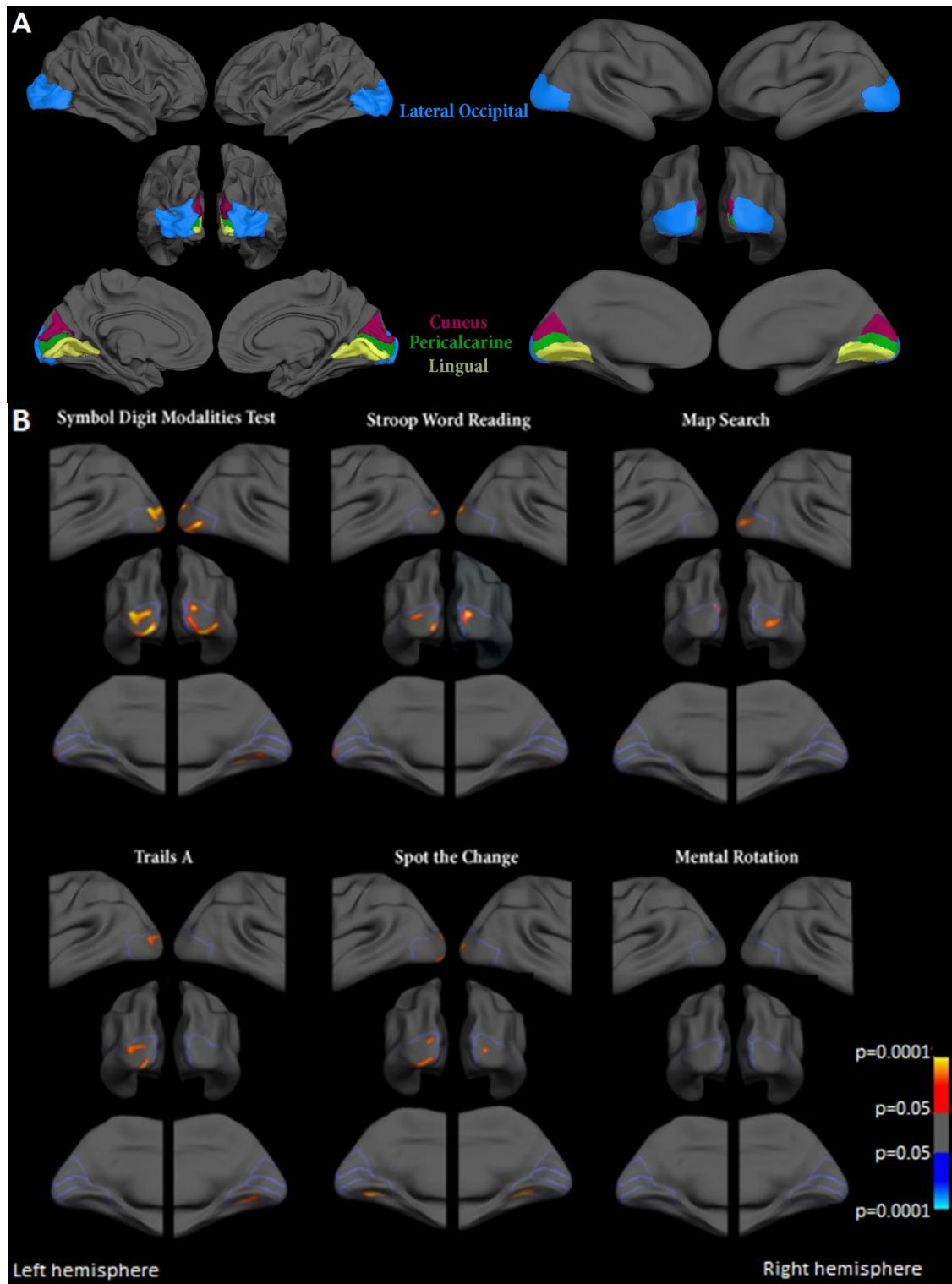


Figure 15-2. A. The LOC, cuneus, lingual and pericalcarine atlas regions as defined by the Desikan-Killiany atlas overlaid on the study-specific average cortical template and on the inflated study-specific average cortical template. B. Significance maps of the associations ($0.0001 < p < 0.05$) between occipital cortex thickness and cognitive task performance displayed on inflated brain templates. Associations were adjusted for age, gender, study site, education, CAG, disease burden score and frontal lobe thickness, and corrected for multiple comparisons using Monte Carlo cluster-wise correction ($p < 0.05$) across the four occipital regions.

15.4 Discussion

This is the first study to show significant, and regionally-specific, associations between occipital cortex thickness and cognitive task performance in an HD cohort with no visual deficits. The associations were most notable within the LOC, but also in the lingual gyrus, which are regions known to be involved in higher-level visual functioning such as word reading, visual memory, integration of visual information and object recognition. In contrast, the regions responsible for more basic visual processing, including the pericalcarine gyrus and cuneus, did not show significant associations with cognitive test performance. Cognitive test performance in diseases such as HD, Alzheimer's disease, Schizophrenia and Bipolar disorder (Holroyd et al. 2000; Lyoo et al. 2006), in which occipital atrophy is present in the absence of overt visual deficits, should be interpreted with the current results in mind; performance is not driven by levels of executive functioning alone but a significant proportion of variance is likely to be due to impairments in higher-level visual processing.

Consistent with previously published findings, HD gene-carriers had thinner occipital cortices both prior to disease onset and in manifest disease, compared with healthy controls (Rosas et al. 2002; Rosas et al. 2005; Rosas et al. 2008; Tabrizi et al. 2009). The between-group differences were heterogeneous across the four occipital regions; thinning of the basic visual processing regions, the pericalcarine and cuneus, produced consistently smaller effect sizes and emerged later in the disease process than thinning of occipital regions involved in higher-level visual processing, i.e. the lingual and LOC. In addition, across the occipital regions the greatest between-group difference appeared between HD gene-carriers nearing onset and those in the early stages of disease (i.e. PreHD-B vs. HD1), which is suggestive of a potential increase in the rate of occipital cortex thinning around disease onset. After disease onset, atrophy appeared to slow somewhat, with no significant difference in any region between gene-carriers with a TFC of 11-13 and those with a TFC of 7-10.

Occipital cortex thickness in HD gene-carriers was associated with performance on the cognitive tasks, all of which included a visual component, but not the Paced Tapping motor task. Paced tapping has been shown to be a significant predictor of disease progression in HD (Tabrizi et al. 2013) and so was included as a general proxy for disease progression with no visual component to test the exclusivity of the visual-occipital relationship. Therefore the lack of associations between Paced Tapping performance level and regional occipital cortex thickness provides evidence that the associations found are specific to visual cognition rather than general disease progression.

The two basic visual processing regions, the pericalcarine and the cuneus, were not associated with cognitive task performance, with the only exception being an association between Stroop Word Reading

performance and cuneus thickness. Conversely, the LOC and lingual regions showed significant relationships with performance on almost all cognitive tests. The thickness of the LOC was estimated to account for between 2.8% and 9.5% of variance in performance on the cognitive tasks. Lingual thickness was estimated to account for 0.3% to 6.1% of performance variance. The LOC is known to, amongst other functions, process object recognition and integrate visual information (Beauchamp 2005; Grill-Spector et al. 2001; Larsson & Heeger 2006) which are highly important aspects of the cognitive tasks examined. The lingual region is known to process colour information (area V4) and visual memory, which may explain the strongest association for this region with Spot the Change performance which requires participants to detect a colour change within an array of items.

Limitations of this study are acknowledged. The HD2 group was smaller than the other groups and therefore the lack of significant progression of cortical thinning from HD1 to HD2 may be a result of reduced power rather than a biological slowing of this atrophy rate. It should also be noted that whilst similar structure-function associations may be seen in Alzheimer's disease, Schizophrenia and Bipolar disorder (Holroyd et al. 2000; Lyoo et al. 2006), in which occipital atrophy is present in the absence of overt visual deficits, the specific pattern of atrophy in HD is likely to differ from that in these other diseases and so further investigations in these clinical cohorts is warranted. As is the case with all cross-sectional studies, these groups contain different people at different disease stages therefore it cannot be assumed that these findings are disease-stage related. Additionally it may be that occipital atrophy is related to another area's atrophy which is the underlying factor behind the cognitive deficits, although based on the hypothesized link between the visual aspects of cognitive task performance and occipital function there is strong reason to believe that this is a direct link.

The significant associations between occipital cortex thickness and cognitive test performance levels demonstrate that while reduced occipital thickness may not manifest itself in obvious visual deficits, there may be substantial effects on higher-level visual functions and consequently cognitive task performance. Overall these results have important implications for the interpretation of cognitive test performance in both HD and other conditions in which occipital atrophy occurs. It is clear that executive function does not account for total cognitive test performance levels; there is a silent contribution of the visual cortex.

16 PART 2: A Summary

Studies within Part 2 of this thesis have led to several advances in our understanding of HD:

Cerebellar GM and WM pathology was shown, for the first time, to have a potential role in the motor and psychiatric symptoms of HD; this study clarified the pattern and impact of these changes. Although not a central region in HD neuropathology this knowledge may help interpret changes or differences in disease phenotype and broadens our knowledge of whole-brain changes in HD.

An exploratory imaging investigation into the macro- and micro-structural imaging correlates of emotion recognition in HD provided evidence for a substantial influence of WM integrity on impaired emotion processing. Results were suggestive of an extensive emotion network differentially recruited by varying emotion processing demands. This was the first study to implicate WM integrity as a significant contributor to this clinical phenotype.

Finally, HD was used as a model of occipital cortex atrophy in the absence of apparent visual deficits in order to investigate the contribution of occipital processing to cognitive task performance. This was the first study to show a significant effect of occipital cortex thickness on task performance levels. The topologically-specific associations were suggestive of higher-level visual impairment in HD, with the relative sparing of regions involved in basic vision. These results take current understanding of performance on tasks of executive function further away from the traditional understanding that these higher-level skills are located exclusively within frontal brain regions (Reitan & Wolfson 1994).

Advances such as these in the understanding and interpretation of clinical and neurological pathology will facilitate future research and good clinical practice, and add to our overall knowledge of the brain and HD. Clinical-imaging associations also strengthen the argument for the clinical relevance of neuroimaging measures as surrogate end-points of disease progression. At this stage of HD research this association has been proven to be valid and now the identification of the strongest biomarkers to assess therapeutic efficacy in HD is a major priority. Therefore the final section of this thesis (Part 3) pulls the findings from Parts 1 and 2 together in a direct statistical comparison of a comprehensive battery of biomarkers in HD; the results from which have been used to produce guidelines for future application in clinical trials.

17 PART 3. Evaluation of Neuroimaging Measures as Biomarkers in HD

Potentially disease-modifying therapeutics are under development in HD (Handley et al. 2006; Mauck et al. 2013; Ross & Tabrizi 2011; Zhang & Friedlander 2011). Therefore there is a requirement for practical guidelines outlining different biomarker options and their power to detect disease progression in different cohorts, over varying intervals. Every clinical trial will require the selection of specific end-points depending on the hypothesized mechanism of action of the drug or therapy, the stage of disease being treated and the design of the study (trial phase, treatment duration, number of patients and study sites, costs etc.).

Assessments of biomarkers in large, multi-site longitudinal observational studies in HD, such as TRACK-HD, PREDICT-HD and most recently the PADDINGTON study, provide evidence as to which measures are reproducible and reliable across multiple sites and most likely to be sensitive to therapeutic intervention. These studies, by design, aimed to replicate the recruitment process, training methods, standardised data collection, blinded QC and centralised independent statistical analysis of clinical trials. Effect sizes are reported for each potential biomarker to quantify each one's sensitivity to HD-related change over time compared with controls. These effect sizes can be transformed to estimated sample size requirements. The relationship between effect sizes and samples sizes for different assumed therapeutic efficacies, with differing levels of statistical power, was shown in Figure 5-1.

The TRACK-HD study (Tabrizi et al. 2012) has provided the broadest list of biomarker effect sizes spanning both pre- to manifest HD to-date (Table 17-1). The largest effect sizes reported in manifest HD were derived from the CBSI, WM, VBSI and BBSI. Of the cognitive measures the SDMT, Indirect Circle Tracing and Stroop Word Reading tasks were the strongest, with roughly similar effect sizes to the clinical measures (TMS and TFC). With an assumed treatment efficacy of 20% in early-stage HD (with 90% power, two-tailed $p < 0.05$) whole-brain volume, TFC and Indirect Circle Tracing biomarkers corresponded to feasible sample size requirements per treatment arm of 419 (95% CI 247, 711), 695 (371, 1618) and 887 (371, 2483) respectively. With an efficacy of 50% these decreased to 68 (95% CI 40, 114), 112 (60, 259) and 142 (60, 398).

There was a much more limited range of potential biomarkers in the TRACK-HD premanifest groups, with cognitive, quantitative motor and psychiatric markers detecting very little change. Several of the neuroimaging measures were however sensitive to change, with TMS also performing strongly. In the preHD-B cohort, with treatment efficacies of 20% and 50% respectively, clinical trials would require, over two years, sample sizes of 193 (95% CI 94, 442) and 31 (15, 71) for caudate atrophy, 353 (170, 727) and 56 (27, 116) for putamen atrophy and 547 (321, 986) and 88 (51, 158) for the TMS.

	PreHD-A	PreHD-B	All PreHD	HD1	HD2	All HD
Neuroimaging measures						
Whole-brain atrophy (% of baseline volume)	0.10 (-0.16 to 0.47)	0.66 (0.32 to 1.05)	0.31 (0.07 to 0.62)	1.12 (0.86 to 1.46)	1.44 (1.00 to 2.46)	1.23 (0.99 to 1.56)
Ventricular expansion (mL)	0.27 (-0.01 to 0.56)	0.90 (0.56 to 1.26)	0.50 (0.29 to 0.77)	1.18 (0.90 to 1.44)	1.03 (0.74 to 1.63)	1.08 (0.86 to 1.38)
Caudate atrophy (% of baseline volume)	0.94 (0.65 to 1.29)	1.65 (1.09 to 2.36)	1.17 (0.92 to 1.47)	1.97 (1.54 to 2.44)	2.18 (1.54 to 3.69)	2.04 (1.68 to 2.48)
Putamen atrophy (% of baseline volume)	0.72 (0.38 to 1.13)	1.22 (0.85 to 1.76)	0.94 (0.67 to 1.25)	1.13 (0.84 to 1.45)	1.03 (0.65 to 1.66)	1.06 (0.80 to 1.38)
White-matter atrophy (% of baseline volume)	0.42 (0.12 to 0.78)	1.04 (0.70 to 1.48)	0.68 (0.45 to 0.94)	1.57 (1.30 to 1.92)	1.91 (1.36 to 3.43)	1.70 (1.40 to 2.08)
Grey-matter atrophy (% of baseline volume)	0.10 (-0.21 to 0.42)	0.44 (0.08 to 0.79)	0.24 (-0.02 to 0.52)	1.01 (0.78 to 1.25)	1.06 (0.78 to 1.41)	1.01 (0.82 to 1.24)
Cognitive measures						
Symbol digit modality test (number correct)	0.03 (-0.25 to 0.34)	0.26 (-0.05 to 0.57)	0.14 (-0.10 to 0.38)	0.88 (0.56 to 1.24)	1.24 (0.81 to 1.59)	1.00 (0.67 to 1.27)
Stroop word reading (number correct)	0.24 (-0.12 to 0.58)	0.05 (-0.23 to 0.47)	0.15 (-0.10 to 0.43)	0.56 (0.31 to 0.91)	1.15 (0.71 to 1.69)	0.73 (0.48 to 1.06)
Emotion recognition (negative emotions, number correct)	-0.19 (-0.46 to 0.25)	0.40 (0.02 to 0.69)	0.08 (-0.18 to 0.38)	0.48 (0.23 to 0.83)	0.49 (0.14 to 1.28)	0.49 (0.27 to 0.85)
Circle tracing—direct (annulus length in log cm)	0.02 (-0.29 to 0.29)	0.26 (-0.07 to 0.65)	0.10 (-0.13 to 0.35)	0.58 (0.31 to 0.90)	0.59 (0.25 to 1.08)	0.58 (0.34 to 0.85)
Circle tracing—indirect (annulus length in log cm)	0.12 (-0.19 to 0.45)	0.29 (-0.03 to 0.75)	0.19 (-0.09 to 0.47)	0.77 (0.46 to 1.19)	1.02 (0.65 to 1.70)	0.85 (0.60 to 1.22)
UHDRS measures						
UHDRS motor assessment (total score)	0.64 (0.44 to 0.99)	0.98 (0.73 to 1.28)	0.79 (0.63 to 0.98)	0.77 (0.54 to 1.05)	0.91 (0.57 to 1.50)	0.81 (0.63 to 1.04)
UHDRS functional capacity* (total score)	0.87 (0.57 to 1.19)	0.83 (0.53 to 1.16)	0.85 (0.62 to 1.09)
Quantitative motor measures						
Chorea position index (log cm/s)	0.20 (-0.09 to 0.49)	0.19 (-0.16 to 0.56)	0.20 (-0.05 to 0.47)	0.61 (0.34 to 0.99)	0.52 (0.18 to 1.02)	0.56 (0.32 to 0.83)
Chorea orientation index (log °/s)	-0.01 (-0.25 to 0.43)	0.03 (-0.32 to 0.33)	0.01 (-0.22 to 0.32)	0.39 (0.09 to 0.69)	0.55 (0.22 to 0.98)	0.45 (0.22 to 0.69)
Grip force variability, 250 g, non-dominant hand (log CV)	-0.02 (-0.34 to 0.33)	0.15 (-0.19 to 0.48)	0.05 (-0.22 to 0.31)	0.41 (0.11 to 0.76)	0.63 (0.28 to 1.10)	0.51 (0.23 to 0.78)
Tongue force variability, 1.5 N target force (log CV)	-0.06 (-0.37 to 0.29)	-0.07 (-0.40 to 0.23)	-0.07 (-0.32 to 0.17)	0.31 (0.02 to 0.64)	0.39 (0.03 to 1.00)	0.33 (0.08 to 0.60)
Speeded-tapping tap duration variability, non-dominant hand (log s)	0.05 (-0.32 to 0.40)	0.32 (-0.03 to 0.67)	0.19 (-0.09 to 0.47)	0.50 (0.17 to 0.79)	1.10 (0.65 to 2.19)	0.68 (0.39 to 1.00)
Speeded-tapping mean inter-tap interval, non-dominant hand (s)	0.35 (0.06 to 0.67)	0.40 (0.11 to 0.72)	0.38 (0.15 to 0.61)	0.26 (0.03 to 0.52)	0.72 (0.41 to 1.05)	0.42 (0.23 to 0.61)
Neuropsychiatric measures						
PBA apathy (square root of points on PBA scale)	0.35 (-0.03 to 0.71)	0.16 (-0.18 to 0.49)	0.27 (0.00 to 0.54)	0.41 (0.13 to 0.72)	0.45 (0.14 to 0.81)	0.41 (0.18 to 0.64)

Table 17-1. 24-month effect sizes (95% CI) from the TRACK-HD study (Tabrizi et al. 2012). All estimates are adjusted for age, sex, education level and study site with the exception of the neuroimaging measures, which were adjusted for age, sex and study site only. CV = coefficient of variation. PBA = problem behaviours assessment. * As controls and premanifest HD groups are not expected to show any change in UHDRS functional capacity, this outcome was modelled and effect sizes are presented for symptomatic HD groups only. When calculating effect sizes, change over 24 months in the HD groups was compared with zero expected change rather than estimated change in controls.

PREDICT-HD (Aylward et al. 2011) has also published effect and sample size estimates for similar structural neuroimaging read-outs in preHD (Table 17-2). This study found that effect sizes in atrophy rates compared

with controls were notably stronger in the mid (9 to 15 YTO) to near (<9 YTO) groups than for the far (>15 YTO) group. This suggests that smaller cohorts would be required if HD gene carriers estimated to be within 15 years of symptom onset were recruited to trials in preHD. In the mid and near onset groups atrophy of the frontal WM, striatum and cerebral WM, along with expansion of ventricular CSF, performed at similar levels in detecting disease-related change. Consequently using any of these biomarkers as end-points would require similar sample sizes. For example, in a two-year clinical trial with a 30% expected therapeutic effect the sample size would need to be between 171 and 524 preHD individuals per treatment arm.

	Far				Mid				Near			
	Effect size difference from controls	Expected therapeutic effect†			Effect size difference from controls	Expected therapeutic effect†			Effect size difference from controls	Expected therapeutic effect†		
		50%	40%	30%		50%	40%	30%		50%	40%	30%
Caudate	−0.33	749	1171	2082	−0.81	126	197	350	−0.60	235	366	652
Putamen	−0.25	1366	2134	3794	−0.60	233	363	646	−0.63	211	329	585
Total striatum	−0.39	535	835	1485	−0.87	111	173	307	−0.75	147	230	409
Thalamus	−0.06	27783	43411	77176	−0.36	661	1032	1835	−0.46	397	620	1103
Globus Pallidus	−0.43	454	710	1262	−0.27	1154	1803	3205	−0.67	188	294	523
Cortical grey	+0.17	3094	4834	8594	+0.22	1739	2717	4829	+0.11	7799	12185	21663
Cerebral white	−0.49	343	535	951	−0.89	106	166	295	−1.17	61	96	171
Total brain	−0.33	748	1169	2078	−0.62	216	338	601	−0.98	87	136	242
Ventricular CSF	+0.31	879	1374	2443	+0.67	188	294	524	+1.20	59	92	163
Surface CSF	−0.11	7697	12027	21381	+0.25	1309	2045	3636	+0.52	303	474	842
Total CSF	+0.05	36241	56627	100671	+0.35	691	1080	1920	+0.69	179	280	497
Frontal white	−0.54	286	447	795	−0.86	112	175	311	−1.15	63	98	174

Table 17-2. PREDICT-HD effect sizes for two-year volume change (Aylward et al. 2011). * N per treatment arm for a 2-year clinical trial, assuming two-sided p=0.05; 90% power. † % reduction in rate of case-control atrophy difference.

A more recent PREDICT-HD publication used data from 1013 preHD individuals (including participants close to onset, up to more than 12.8 years from estimated onset) and 301 control participants over a 10-year interval (Paulsen et al. 2014). In contrast to the TRACK-HD and earlier PREDICT-HD analyses which input change data directly into their models, this study averaged data within each group at each annual time-point and computed change slopes from these group averages. The consequent estimated sample size requirements are surprising small (Table 17-3).

Variable	Type	Dropout (%)	Estimated parameters					Effect size						
			β_s	g_{11}	g_{12}	g_{22}	σ_e^2	20%	30%	40%	50%	60%	70%	
CSF	Imaging	0	4.094	486.882	5.189	3.183	53.737	332	147	83	53	37	27	
CSF	Imaging	10	4.094	486.882	5.189	3.183	53.737	357	159	89	57	40	29	
CSF	Imaging	20	4.094	486.882	5.189	3.183	53.737	386	172	97	62	43	32	
Putamen	Imaging	0	−0.100	0.810	−0.004	0.003	0.037	391	174	98	63	43	32	
Putamen	Imaging	10	−0.100	0.810	−0.004	0.003	0.037	421	187	105	67	47	34	
Putamen	Imaging	20	−0.100	0.810	−0.004	0.003	0.037	455	202	114	73	51	37	
Caudate	Imaging	0	−0.111	0.666	0.005	0.003	0.078	628	279	157	101	70	51	
Caudate	Imaging	10	−0.111	0.666	0.005	0.003	0.078	676	301	169	108	75	55	
Caudate	Imaging	20	−0.111	0.666	0.005	0.003	0.078	732	325	183	117	81	60	
TMS	Motor	0	1.498	21.338	5.008	3.050	16.789	981	436	245	157	109	80	
TMS	Motor	10	1.498	21.338	5.008	3.050	16.789	1051	467	263	168	117	86	
TMS	Motor	20	1.498	21.338	5.008	3.050	16.789	1131	503	283	181	126	92	
Sp-Tapping	Cognitive	0	7.345	3255.107	296.690	72.380	554.870	1230	546	307	197	137	100	
Sp-Tapping	Cognitive	10	7.345	3255.107	296.690	72.380	554.870	1318	586	330	211	146	108	
Sp-Tapping	Cognitive	20	7.345	3255.107	296.690	72.380	554.870	1421	632	355	227	158	116	
Hippocampus	Imaging	0	−0.022	0.066	0.000	0.000	0.006	1322	588	330	212	147	108	
Hippocampus	Imaging	10	−0.022	0.066	0.000	0.000	0.006	1422	632	356	228	158	116	
Hippocampus	Imaging	20	−0.022	0.066	0.000	0.000	0.006	1540	684	385	246	171	126	
Chorea	Motor	0	0.480	1.383	0.401	0.335	2.675	1373	610	343	220	153	112	
Chorea	Motor	10	0.480	1.383	0.401	0.335	2.675	1472	654	368	236	164	120	
Chorea	Motor	20	0.480	1.383	0.401	0.335	2.675	1586	705	397	254	176	130	
Brady	Motor	0	0.520	4.748	0.811	0.438	3.305	1467	652	367	235	163	120	
Brady	Motor	10	0.520	4.748	0.811	0.438	3.305	1574	699	393	252	175	128	
Brady	Motor	20	0.520	4.748	0.811	0.438	3.305	1698	754	424	272	189	139	
Dysrhythmia	Cognitive	0	3.002	592.124	53.942	12.919	135.291	1676	745	419	268	186	137	
Dysrhythmia	Cognitive	10	3.002	592.124	53.942	12.919	135.291	1801	800	450	288	200	147	
Dysrhythmia	Cognitive	20	3.002	592.124	53.942	12.919	135.291	1945	864	486	311	216	159	
Thalamus	Imaging	0	−0.058	0.316	−0.006	0.003	0.069	2077	923	519	332	231	170	
Thalamus	Imaging	10	−0.058	0.316	−0.006	0.003	0.069	2230	991	557	357	248	182	
Thalamus	Imaging	20	−0.058	0.316	−0.006	0.003	0.069	2407	1070	602	385	267	197	
TFC	Functional	0	−0.156	0.518	0.038	0.054	0.551	2540	1129	635	406	282	207	
TFC	Functional	10	−0.156	0.518	0.038	0.054	0.551	2722	1210	680	435	302	222	
TFC	Functional	20	−0.156	0.518	0.038	0.054	0.551	2932	1303	733	469	326	239	
SDMT	Cognitive	0	−0.906	124.336	0.742	0.904	20.943	2547	1132	637	408	283	208	
SDMT	Cognitive	10	−0.906	124.336	0.742	0.904	20.943	2740	1218	685	438	304	224	
SDMT	Cognitive	20	−0.906	124.336	0.742	0.904	20.943	2964	1317	741	474	329	242	

Table 17-3. PREDICT-HD 10-year data (Paulsen et al. 2014): Estimated sample size requirements (right side) for a two-arm phase II randomised clinical trial. Effect size is the percentage difference in rate of change (slope) of the treated and untreated groups. SP-Tapping, speeded tapping; Brady, bradykinesia.

Inconsistencies between these study effect sizes and sample size estimates may be influenced by cohort and/or methodological differences. Imaging analysis in the PREDICT-HD study was fully automated (BRAINS software) whilst TRACK-HD applied more manual intervention. Therefore, even with identical statistical analysis methods, the pattern of sensitivity between different regions may still differ depending on the strengths of each technique.

One limitation of the current guidelines is that effect sizes are derived from two- and ten-year data. Clinical trials, especially in their early stages, require read-outs from much shorter intervals. The choice of HD cohort most likely to be recruited to clinical trials must also be considered. Several neuroimaging biomarkers perform strongly in preHD groups but these individuals are by definition symptom-free and therefore detecting a direct, clinical improvement from a disease-modifying therapeutic will be very difficult. Stage I patients, the very early clinical stages of this neurodegenerative disease, are the symptomatic group most likely to be recruited for imminent clinical trials. In this group functional impairment is minimal and benefit is most likely to be realised and detectable. The PADDINGTON study (WP2) was specifically designed to match these criteria; a comprehensive comparison between cognitive, clinical and neuroimaging biomarkers over short and varying time intervals was conducted in early-stage HD, providing practical guidelines for biomarker selection in future clinical trials.

18 Evaluation of Multi-Modal, Multi-Site Neuroimaging Measures as Biomarkers in HD: Results from the PADDINGTON study

18.1 Background

The PADDINGTON study builds on TRACK-HD and PREDICT-HD publications which have provided estimates of sample size requirements for clinical trials with different biomarkers as outcome measures. Clinical trials require read-outs over shorter time intervals than previously tested, therefore the PADDINGTON study evaluation of cognitive, clinical and neuroimaging biomarkers was conducted over intervals of six, nine and 15 months. The biomarker battery was also adapted to include, for the first time, diffusion (micro-structural) neuroimaging measures, lobular cortical thickness estimates and short interval VBM.

18.1.1 Micro- versus Macro-Structural ROI-Based Biomarkers

Although both micro- and macro-structural neuroimaging metrics are proposed as biomarkers for HD (Dumas et al. 2012; Paulsen et al. 2008; Rosas et al. 2006; Tabrizi et al. 2012; Weaver et al. 2009), there is little evidence directly comparing the two modalities in terms of their relative sensitivities. For inherent biological reasons, it is assumed that micro-structural measures may show improved sensitivity to HD-related pathology compared with macro-structural measures, i.e. it would be expected that disruption of cellular membranes and axonal injury would precede gross morphometric changes. However, one small study comparing striatal volume with striatal Trace values suggested the opposite (Vandenberghe et al. 2009). It is difficult to draw conclusions from this early study since numbers were small (10 HD and 12 control participants), the acquisition protocol was basic (three directions for the diffusion-weighted sequence) and there was no formal statistical comparison of modalities. Conversely, another study using a voxel-wise approach to distinguish preHD participants from controls, found that diffusion imaging gave a better separation than T1-weighted imaging (Kloppel et al. 2008). Hence there is a need to perform a robust statistical evaluation of the relative sensitivities of macro- and micro- structural neuroimaging measures in HD.

18.1.2 Cortical Thickness

Cortical thinning as a biomarker for clinical trials in HD is most pertinent for therapeutics that are hypothesized to have a beneficial impact on the cortex. Results from the first use of FreeSurfer software's cortical thickness analysis as an end-point in a clinical trial in HD was recently published as part of the PRECREST study (Rosas et al. 2014). To-date however there has been no direct comparison of cortical thickness as a biomarker against other more widely published neuroimaging measures in HD. Based on the results of preliminary cortical thickness analyses and software tests (Chapter 9) a parcellation approach per lobe was suggested as the optimal use of cortical thickness data as a biomarker; extracting average

thickness data from either the parietal or occipital lobe. These metrics will be included in the following comparison.

18.1.3 Short-Interval VBM

VBM is a widely used tool, in SPM software, for assessing brain changes across the whole-brain at the voxel-level. Combined with fluid-registration it can be used to create SPMs of longitudinal volumetric neurodegenerative changes. As such it is a potential neuroimaging biomarker in HD. This whole-brain coverage can be argued to be an advantage over ROI analyses as it allows for detection of atrophy across larger areas and a variety of brain regions. To-date the shortest VBM analysis reporting SPMs in HD has been over a 12-month interval (Tabrizi et al. 2011); detecting widespread WM atrophy. Inclusion of VBM analyses in this method comparison will test whether it is sensitive to HD-related changes over shorter scan intervals; six and nine months. This will allow assessment of whether VBM SPMs are appropriate as short-interval neuroimaging biomarkers in HD and whether this mass-univariate approach may add valuable supplementary information to hypothesis-driven ROI volumetric biomarkers.

In summary, this study is the first comprehensive comparison of cognitive, clinical and neuroimaging biomarkers to be conducted over clinically relevant, short time intervals in early-stage HD. It will also provide the first statistical evaluation in HD of: micro- versus macro-structural imaging measures; cortical thickness as a biomarker in direct comparison with other neuroimaging measures; and VBM over short scan intervals. Macro- and micro-structural neuroimaging measures, including cortical thickness analysis, will be evaluated using a hypothesis-driven approach, examining metrics over ROIs previously implicated in HD. These metrics will be contrasted with a hypothesis-free mass univariate VBM analysis. Direct statistical comparison will determine whether any of the cognitive or clinical tests, brain regions or imaging modalities investigated confer a significant advantage over the others in terms of sensitivity to HD pathology.

18.2 Methods

18.2.1 Cohort

A total of 40 controls and 61 early HD participants were recruited as part of the PADDINGTON study. All participants completed assessments at baseline. 40/40 controls and 57/61 HD participants completed the 6-month visit. 37/40 controls and 56/61 HD participants returned for the 15-month visit. 37/40 controls and 53/61 HD participants have data from all three time-points. Only one HD participant did not return for the 6- or 15-month visits and therefore no longitudinal data was available from this subject for inclusion in this method comparison.

18.2.2 Image Acquisition and Analysis

Participants were scanned at baseline, six and 15 months following the acquisition protocols detailed in Section 3.4. Image pre-processing and details of the analysis protocol are described in Sections 4.2 and 4.3.

MRI: Macro-Structural (Volumetric) Analysis

The software package MIDAS (Freeborough et al. 1997) was used to delineate the whole-brain, caudate, CC and ventricles at baseline (Section 4.2.2). BRAINS3 software (Magnotta et al. 2002) was chosen for putamen analysis over manual delineation due to low signal with the manual method and previously proven longitudinal sensitivity of this automated technique in TRACK-HD (Tabrizi et al. 2011). Change in whole-brain, caudate and ventricular volume over the scanning interval was estimated using the BSI technique (Freeborough & Fox 1997). Change in CC and putamen volume was estimated by delineating the structures at both time-points and subtracting the volumes at each time-point. GM and WM volume changes were computed using a fluid-registration approach (Christensen et al. 1996; Hobbs et al. 2010b; Tabrizi et al. 2011) – details in Section 4.2.10. Briefly, fluid registration was applied to align the follow-up image to the baseline (Christensen et al. 1996). Logged Jacobian determinants from this registration were then masked by SPM-derived GM and WM tissue segmentations and quantified to output an estimate of volume change within these regions. FreeSurfer (version 5.3.0) was used to compute thickness estimates for the occipital and parietal lobes; the rationale and pipeline for this analysis was described in full in Chapter 9.

Whole-brain longitudinal VBM was also performed in SPM8; details of this analysis were previously described in Section 4.2.9. Briefly, Unified Segmentation was run to extract GM and WM tissue segmentations from the baseline image in native-space. The follow-up image was fluidly registered to the baseline (Christensen et al. 1996). The logged Jacobian determinants of this registration were reoriented to match the SPM8-format baseline scans and masked by the GM and WM tissue segmentations. These masked regions were then warped onto a study-specific DARTEL template (Ashburner 2007) and smoothed using a 4mm FWHM kernel. Statistical analyses were conducted on every voxel of these smoothed, normalised, masked logged determinants resulting in SPMs which were adjusted for multiple comparisons using FWE correction at the $p < 0.05$ level.

MRI: Micro-Structural (Diffusion) Analysis

Camino software (<http://www.cs.ucl.ac.uk/research/medic/camino>) was used to generate diffusion images. FA, MD, RD and AD were computed over four ROIs; the WM, CC, caudate and putamen. These regions were defined on the T1-weighted images for each individual, eroded by one voxel and saved as binary masks. The eroded ROIs were then transformed into 'FA-space' using NiftyReg (<http://sourceforge.net/projects/niftyreg> (Modat et al. 2010; Ourselin et al. 2001)). The transformations

generated during these registrations were then applied to the eroded ROIs and averaged diffusion metrics were calculated over these regions using the *fsstats* utility within the FSL toolbox (Smith et al. 2004).

For the longitudinal analysis, in order to avoid asymmetrical registration bias a common longitudinal mask was generated for each region by defining a subject-specific 'half-way-space'. The baseline T1 scan was affine registered (method adapted from Smith et al. (Smith et al. 2001)) to the follow-up image and vice versa and then the symmetric 'average' transformation was calculated between the two time-points. Using these two symmetric transformations, a mid-point (i.e. the half-way-space) was defined and then transformation to this space applied to native-space images and their corresponding regional masks. Any voxels not present in the ROI at both time-points were removed to only retain those common to both baseline and follow-up. B0, FA and diffusion maps for each time-point were also transformed into this space. Using NiftyReg, the half-way T1 images were then warped into half-way FA-space (Modat et al. 2010;Ourselin et al. 2001). Finally, using the inverse of the B0 native-to-half-way-space transformation, the T1 images and half-way-space common mask regions were moved into native diffusion space for each of the baseline and follow-up scans. Regional means for FA, MD, RD and AD were generated within these regions using the FSL utility, *fsstats*.

Quality Control (QC)

Raw T1-weighted scans were visually inspected for quality. Two T1 scans at the 6-month visit and one at the 15-month visit failed QC (all HD participants). All manual delineations and BRAINS3 segmentations were visually inspected, blind to diagnosis. BSI affine registrations and fluid registration VCMs were assessed for quality in MIDAS software. Those deemed to be of insufficient quality were removed from further analyses. Of the six, nine and 15-month affine registrations, 11/95, 11/87 and 12/92 failed QC (45.5%, 45.5% and 58.3% of which were HD scans respectively). Of the six and nine month fluid registrations for the VBM analysis, 16/95 and 11/87 failed QC (50% and 45.5% of which were HD scans respectively). SPM tissue segmentations deemed to be incomplete, blurred or flawed were rerun after reorientation of the original scan to a more central, upright position in the FOV. Those failing QC a second time were removed from further analyses. This included five scans at baseline (2 control and 3 HD scans) and five from the 6-month visit scans (2 control and 3 HD scans). The VBM analysis was additionally run with the reverse contrast. This confirmed that no regions were detected with faster atrophy rates in the control group compared with the HD group in either the GM or WM.

For the diffusion analysis all regions and registrations were visually inspected for quality, both in T1- and FA- space. Four participants were missing data-points for all diffusion metrics: two due to severe motion artefacts and two due to poor T1-to-FA registration. A total of seven WM region diffusion metrics are

missing as three further participants had occipital regions positioned outside the FOV as a result of poor head positioning.

18.2.3 Cognitive and Clinical Assessments

A comprehensive battery of cognitive and clinical assessments was also assessed as part of the PADDINGTON study visit by clinicians and raters trained in the delivery of each test/measurement.

18.2.4 Statistical Analysis

Statistical analysis was performed according to a pre-defined analysis plan. Each outcome was separately analysed for between-group differences using generalised least squares regression models, adjusting for age, gender, study site and scan interval (longitudinal analyses only). Models for between-group differences in cognitive and clinical measures were additionally adjusted for education level (ISCED). Estimates are reported with bias-corrected 95% CIs. Effect sizes were calculated, as previously specified (Section 5.3), as the adjusted difference in the mean of the metric between groups divided by the estimated residual SD of the HD group. All effect sizes have been inverted such that a positive effect size suggests between-group differences in the 'expected' direction.

A linear mixed model was used to compute effect sizes over 6-, 9- and 15-month intervals for all biomarkers under evaluation. All three measurements (6-, 9- and 15-month change) were included in one model that estimated the between-group difference in change over three months - this allowed for non-exact intervals and acceleration effects over time. Combinations of this change and a quadratic term were used to calculate the change and variance for each of the three intervals. Due to known additivity issues with GM and WM change metrics, and unpublished TRACK-HD study data which shows that this method does not fully capture change over periods longer than 12 months, these two variables were modelled using only the 6- and 9-month change data.

When each statistical comparison being made is of independent scientific interest there is a good argument for not making any adjustment to p-values for the fact that multiple comparisons have been made. This was believed to be the case when comparing the effect sizes here and consequently this was the policy adopted.

Between-group differences in tissue volume change assessed using VBM were computed using a factorial design adjusting for age, gender, site and scan interval. Results were adjusted for multiple comparisons using FWE correction at the $p < 0.05$ level.

18.3 Results

Figure 18-1 presents the estimated cross-sectional effect sizes and 95% CIs for each of the 21 outcomes, grouped by imaging metric and modality. The two largest effect sizes were observed for the macro-

structural metrics of putamen volume and caudate volume, with estimates of 2.41 (95% CI 1.75, 2.94) and 2.35 (1.58, 2.96) respectively. These effect sizes were significantly larger than for all of the other metrics ($p < 0.05$), with the exception of axial-, radial- and mean- diffusivity in the putamen. The other macro-structural outcomes of whole-brain, lateral ventricular and CC volume showed comparatively smaller effect sizes of approximate similarity. Structure-specific diffusivity effect sizes were broadly similar irrespective of 'direction', i.e. radial, axial or mean. Diffusivity effect sizes were largest for the putamen, followed by the caudate. The FA effect sizes were smaller in magnitude than both the volume and diffusivity metrics for the corresponding structure.

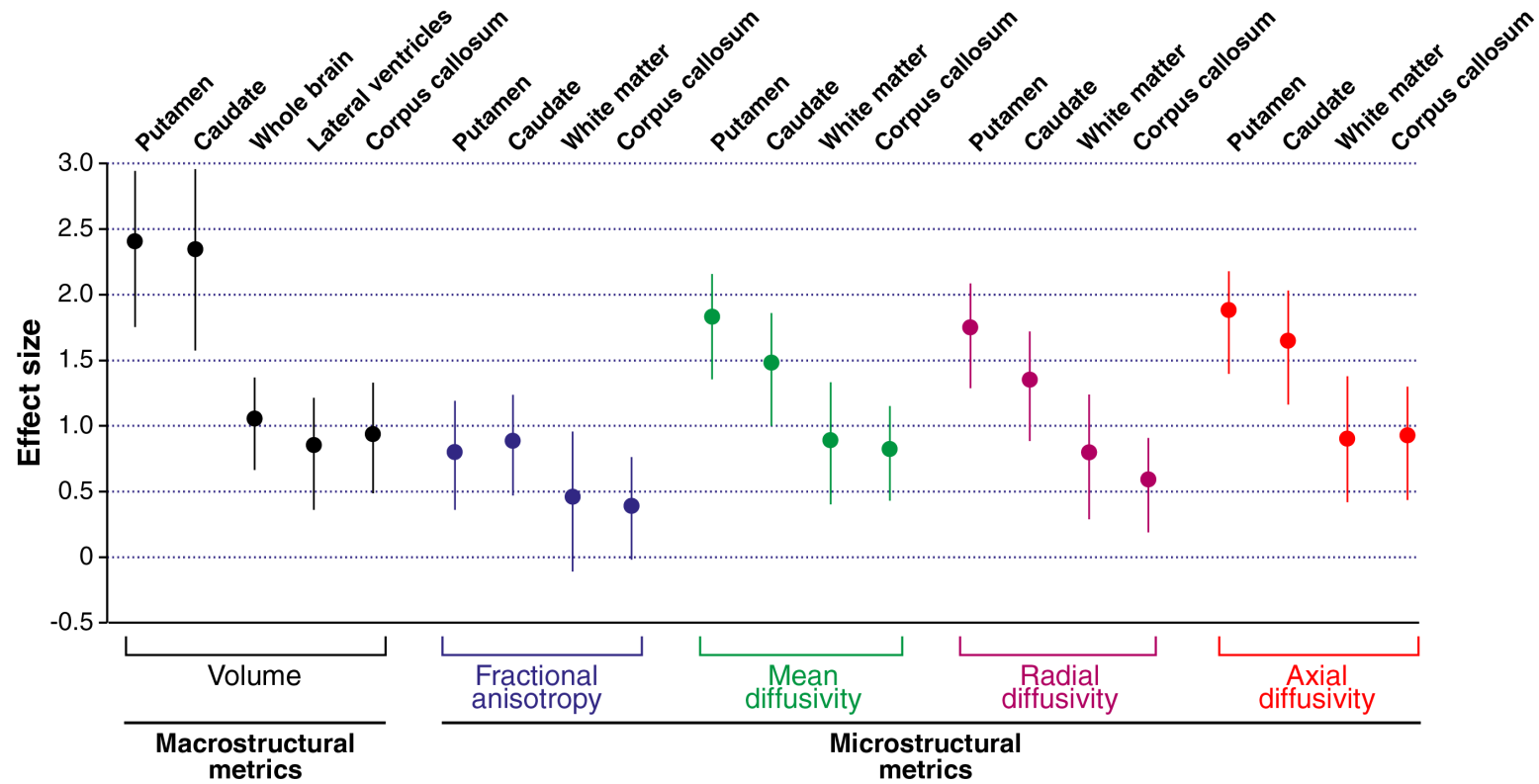


Figure 18-1. Cross-sectional effect size estimates (mean and 95% CIs) for each of the imaging outcomes. Data are grouped by imaging metric and modality. Effect sizes were calculated as the absolute difference in the mean of the metric between groups, adjusted for age, gender and study site, divided by the estimated residual SD of the HD group.

Group averages of the indirect measures of change are reported at baseline, 6- and 15-month visits in Table 18-1, along with between-group differences in change over the three (6-, 9- and 15-month) intervals. Over the 6-month interval only the diffusion measures, HVLT delayed recall and SDMT detected statistically significant between-group differences. Over 15 months, almost all the diffusion measures, putamen volume, occipital cortex thickness, TMS and four of the 11 cognitive tests detected significant between-groups differences. Improved performance was seen in several of the cognitive measures in the control group but not the HD group. These improvements contributed to the significant between-group differences in performance over time in the Letter Fluency, HVLT delayed recall, SDMT and Stroop Interference tasks. As with the cross-sectional data, diffusivity metrics (MD, RD and AD) detected notably more significant differences than the FA measures, except in the CC where FA was more sensitive to longitudinal change. Of the cortical thickness measures only occipital lobe atrophy showed a significant between-group difference and this was only over the longer 15-month interval; -0.0235mm (95% CI $-0.045, -0.021$) $p=0.032$. Direct measures of change are reported in Table 18-2. All BSI metrics, along with GM and WM fluid-based atrophy measures, detected highly significant between-group differences in atrophy rates over all scan intervals tested.

Over the 6-month interval, cognitive effect sizes ranged from 0.637 for the SDMT to 0.114 for the Trails B task (Table 18-3). Of these, only Letter Fluency showed a significant between-group difference over 9 months with an effect size of 0.618. This test was also strong over 15 months (0.664) although not as strong as the SDMT (0.799). Of the two clinical scales, TMS out-performed TFC in its sensitivity to change over time in this early-stage HD cohort. The TMS effect sizes over six, nine and 15 months (0.047, 0.58 and 0.545 respectively) were however not among the strongest within the battery of biomarkers tested. Micro-structural neuroimaging measures performed relatively strongly. Diffusivity measures (MD, AD and RD) proved more sensitive than FA metrics in the GM of the caudate and putamen but not the WM of the CC. The effect sizes of caudate MD over the intervals tested (0.539, 0.62 and 1.111 respectively) are comparable to some of the strongest macro-structural regional biomarkers. Overall, the macro-structural CBSI, VBSI and WM atrophy measures were the top performing biomarkers, with good effect sizes over as short an interval as six months (0.697, 0.791 and 0.615 respectively). The cortical thickness estimates were the weakest of all the biomarker modalities, with occipital lobe atrophy showing an effect size of just 0.512 over 15 months.

Table 18-1. Biomarker group averages at baseline, 6- and 15-month visits, with adjusted between-group differences in change estimates.

Measure	Baseline		6 months		15 months		Adjusted Between-Group Differences in Change Estimates (95% CI) and P-Values					
	Control	HD	Control	HD	Control	HD	6-month Interval		9-month Interval		15-month Interval	
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Estimate (95% CI)	P-Value	Estimate (95% CI)	P-Value	Estimate (95% CI)	P-Value
Cognitive Measures												
Letter Fluency	42.82 (11.15)	28.7 (13.22)	45.73 (11.68)	30.92 (14.19)	46.57 (12.36)	28.91 (13.8)	-0.8 (-3.49, 1.89)	0.561	-3.66 (-6.89, -0.43)	0.027	-4.46 (-7.99, -0.92)	0.013
Category Fluency	23.8 (4.99)	16.84 (6.13)	23.85 (4.27)	15.83 (6.47)	23.35 (4.36)	15.43 (6.11)	-0.98 (-2.76, 0.8)	0.28	-0.47 (-2.34, 1.4)	0.621	-1.45 (-3.41, 0.5)	0.144
HVLT delayed recall (/12)	10.03 (1.54)	6.95 (2.96)	10.2 (2)	6.29 (3.35)	10.27 (2.05)	6.38 (3.36)	-0.82 (-1.53, -0.11)	0.023	-0.002 (-0.85, 0.845)	0.996	-0.82 (-1.67, 0.02)	0.057
HVLT total correct (/12)	10.85 (1.23)	9.15 (2.57)	10.98 (1.12)	8.46 (3.5)	10.89 (2.75)	9.38 (3)	-0.44 (-2.05, 1.34)	>0.05	-0.50 (-2.56, 1.27)	>0.05	-0.95 (-2.85, 0.68)	>0.05
HVLT recognition (/36)	28.95 (3.66)	21.87 (6.39)	28.7 (4.48)	21.05 (6.31)	28.92 (3.67)	21.09 (6.67)	-0.61 (-1.66, 0.20)	>0.05	0.983 (-0.002, 2.542)	>0.05	0.37 (-0.70, 1.86)	>0.05
Symbol Digit Modalities	53.68 (9.02)	36.52 (13.04)	55.73 (10.09)	35.58 (13.17)	56.05 (9.92)	35.36 (13.48)	-2.68 (-4.56, -0.8)	0.005	-1.87 (-4.05, 0.32)	0.094	-4.55 (-6.81, -2.29)	<0.001
Trail A Time (seconds)	24.9 (8.78)	44.11 (24.25)	25.55 (8.61)	48.64 (29.67)	23.76 (8.25)	43.29 (17.28)	3.21 (-1.06, 7.76)	>0.05	-0.68 (-5.21, 2.96)	>0.05	2.52 (-1.36, 6.51)	>0.05
Trail B Time (seconds)	59.9 (24.88)	120.12 (66.27)	61.6 (38.78)	128.64 (69.67)	66.05 (38.92)	121.43 (63.74)	4.36 (-8.64, 17.42)	>0.05	-7.41 (-21.13, 4.77)	>0.05	-3.05 (-17.87, 10.50)	>0.05
Stroop Word	108.55 (15.29)	78.46 (19.02)	107.23 (16.75)	75.08 (19.79)	108.22 (15.39)	74.41 (20.33)	-2.37 (-5.16, 0.42)	0.096	-0.53 (-3.42, 2.36)	0.721	-2.9 (-6.41, 0.61)	0.106
Stroop Colour	82.1 (12.35)	60.39 (14.62)	81.7 (12.5)	57.36 (16.64)	83.3 (12.01)	58.73 (16.2)	-2 (-5.37, 1.37)	0.245	-2.03 (-6.02, 1.96)	0.319	-4.03 (-7.66, -0.39)	0.03
Stroop Interference	46.9 (9.8)	34.23 (11.42)	48.35 (9.16)	34.2 (11.45)	48.54 (9.15)	33.75 (12)	-1.13 (-3.4, 1.14)	0.33	-1.71 (-3.75, 0.32)	0.098	-2.84 (-5.21, -0.47)	0.019
Clinical Measures												
TMS (square root)	0.79 (0.88)	4.33 (1.15)	0.84 (0.86)	4.28 (1.24)	0.61 (0.8)	4.43 (1.47)	0.03 (-0.28, 0.34)	0.85	0.4 (0.08, 0.71)	0.013	0.42 (0.09, 0.76)	0.014
TFC score (/13) *	12.98 (0.16)	11.74 (1.45)	13 (0)	11.36 (1.78)	13 (0)	11.52 (1.51)	-0.31 (-1.18, 0.35)	>0.05	-0.18 (-1.37, 0.93)	>0.05	-0.49 (-1.78, 0.47)	>0.05
Micro-Structural Neuroimaging Measures (x10⁻⁵)												
Caudate FA	212.94 (32.07)	238.99 (46.48)	213.04 (35.12)	238.73 (49.52)	218.21 (30.87)	233.51 (46.35)	-65.97 (-142.55, 10.61)	0.091	-66.67 (-147.41, 14.07)	0.106	-132.64 (-219.81, -45.47)	0.003
Caudate MD (mm ² /s)	7.74 (0.94)	8.62 (1.34)	7.78 (0.96)	8.93 (1.82)	7.67 (0.78)	9.34 (1.77)	0.42 (0.17, 0.67)	0.001	0.52 (0.26, 0.78)	<0.001	0.95 (0.67, 1.22)	<0.001
Caudate RD (mm ² /s)	7.01 (0.9)	7.74 (1.27)	7.05 (0.94)	8.03 (1.75)	6.93 (0.76)	8.44 (1.71)	0.41 (0.16, 0.66)	0.001	0.51 (0.25, 0.78)	<0.001	0.92 (0.65, 1.19)	<0.001
Caudate AD (mm ² /s)	9.19 (1.03)	10.39 (1.52)	9.23 (1.03)	10.72 (2)	9.14 (0.84)	11.14 (1.92)	0.46 (0.2, 0.71)	0.001	0.54 (0.27, 0.8)	<0.001	0.99 (0.72, 1.26)	<0.001
Putamen FA	207.08 (38.38)	231.65 (49.42)	206.17 (41.88)	236.66 (50.46)	205.75 (39.3)	236.09 (51.42)	7.72 (-57.93, 73.37)	0.818	42.44 (-22.62, 107.5)	0.201	50.15 (-16.44, 116.75)	0.14
Putamen MD (mm ² /s)	7.03 (0.38)	8.11 (0.76)	7 (0.37)	8.06 (0.74)	6.98 (0.34)	8.17 (0.89)	0.12 (0.04, 0.2)	0.003	0.078 (0.002, 0.155)	0.045	0.2 (0.12, 0.28)	<0.001
Putamen RD (mm ² /s)	6.32 (0.42)	7.2 (0.67)	6.29 (0.43)	7.13 (0.67)	6.28 (0.38)	7.23 (0.84)	0.1 (0.02, 0.18)	0.021	0.06 (-0.02, 0.15)	0.127	0.16 (0.08, 0.25)	<0.001

Putamen AD (mm ² /s)	8.46 (0.41)	9.94 (1.03)	8.42 (0.35)	9.91 (1.01)	8.4 (0.38)	10.04 (1.1)	0.17 (0.08, 0.26)	<0.001	0.11 (0.02, 0.19)	0.012	0.28 (0.18, 0.37)	<0.001
WM FA	496.58 (21.56)	485.24 (23.67)	498.22 (23.4)	486.49 (21.28)	495.29 (23.38)	484.45 (22.52)	-20.28 (-47.94, 7.39)	0.151	7.21 (-22.19, 36.61)	0.631	-13.07 (-47.34, 21.2)	0.455
WM MD (mm ² /s)	7.26 (0.23)	7.47 (0.24)	7.24 (0.27)	7.51 (0.25)	7.25 (0.26)	7.52 (0.25)	0.04 (0.01, 0.08)	0.013	0.02 (-0.02, 0.05)	0.274	0.06 (0.02, 0.1)	0.002
WM RD (mm ² /s)	5.1 (0.26)	5.32 (0.29)	5.08 (0.3)	5.34 (0.29)	5.1 (0.3)	5.35 (0.29)	0.04 (0, 0.07)	0.034	0.01 (-0.02, 0.04)	0.527	0.05 (0.01, 0.09)	0.011
WM AD (mm ² /s)	11.59 (0.25)	11.78 (0.22)	11.56 (0.27)	11.85 (0.24)	11.54 (0.27)	11.85 (0.22)	0.06 (0.01, 0.11)	0.012	0.04 (-0.02, 0.09)	0.19	0.09 (0.03, 0.16)	0.004
CC FA	735.47 (51.55)	724.37 (59.85)	735.65 (54.06)	719.89 (58.25)	737.33 (54.15)	714.6 (56.26)	-105.23 (-186.79, -23.66)	0.011	-40.99 (-124.96, 42.98)	0.339	-146.21 (-233.66, -58.76)	0.001
CC MD (mm ² /s)	7.76 (0.77)	8.11 (0.86)	7.77 (0.89)	8.19 (0.88)	7.55 (0.73)	8.03 (0.66)	0.08 (-0.05, 0.22)	0.231	0.06 (-0.09, 0.21)	0.409	0.15 (-0.03, 0.32)	0.1
CC RD (mm ² /s)	3.67 (0.8)	3.93 (0.94)	3.69 (0.89)	4.02 (0.96)	3.55 (0.77)	3.96 (0.78)	0.13 (-0.01, 0.27)	0.062	0.05 (-0.1, 0.19)	0.542	0.18 (0.01, 0.34)	0.035
CC AD (mm ² /s)	15.93 (0.98)	16.46 (1.01)	15.94 (1.11)	16.51 (1.04)	15.56 (0.9)	16.17 (0.77)	0.01 (-0.18, 0.19)	0.935	0.13 (-0.07, 0.33)	0.206	0.14 (-0.09, 0.37)	0.242
Macro-Structural Neuroimaging Measures (ml as % TIV; % Change from baseline)												
Putamen	0.56 (0.06)	0.38 (0.08)	0.56 (0.06)	0.38 (0.08)	0.57 (0.06)	0.37 (0.08)	-0.538 (-2.086, 1.011)	0.496	-3.02 (-4.513, -1.526)	<0.001	-3.56 (-5.16, -1.95)	<0.001
CC	0.374 (0.052)	0.323 (0.063)	0.372 (0.051)	0.321 (0.06)	0.378 (0.052)	0.327 (0.053)	-0.19 (-0.84, 0.42)	>0.05	-0.6 (-1.75, 0.61)	>0.05	-0.8 (-2.07, 0.54)	>0.05
Cortical Thickness Measures (mm ×10 ⁻²)												
Occipital Lobe	1.86 (0.01)	1.76 (0.12)	1.86 (0.10)	1.76 (0.12)	1.87 (0.10)	1.74 (0.12)	-1.27 (-3.21, 0.68)	0.201	-5.54 (-13.55, 2.48)	0.176	-2.35 (-4.5, -0.21)	0.032
Parietal Lobe	2.31 (0.10)	2.20 (0.14)	2.30 (0.11)	2.2 (0.12)	2.29 (0.12)	2.18 (0.15)	-0.25 (-2.35, 1.85)	0.816	-1.4 (-9.94, 7.13)	0.747	-2.04 (-4.48, 0.4)	0.101
Generalised least squares regression models were fitted, adjusting for age, gender, study site, scan interval and education level (cognitive and clinical only) and reported with bias-corrected 95% CIs. Grey highlights significant (p<0.05) between-group differences. * TFC change relative to zero in HD group only.												

Table 18-2. 6-, 9- and 15-month regional BSI and fluid-based estimates of change with adjusted between-group differences and p-values.

Measure	Atrophy Rates						Adjusted Between-Group Difference Estimates (95% CI) and P-values					
	6-Month Atrophy (% baseline, vents in ml)		9-Month Atrophy (% baseline, vents in ml)		15-Month Atrophy (% baseline, vents in ml)		6-Month Interval		9-Month Interval		15-Month Interval	
	Controls	HD	Controls	HD	Controls	HD	Estimate (95% CI)	P-Value	Estimate (95% CI)	P-Value	Estimate (95% CI)	P-Value
CBSI	0.572 (1.277)	2.033 (1.835)	0.622 (1.420)	1.867 (2.488)	0.971 (1.314)	3.832 (2.583)	1.38 (0.763 , 1.997)	<0.001	1.723 (0.971 , 2.475)	<0.001	3.103 (2.328, 3.878)	<0.001
BBSI	0.151 (0.418)	0.519 (0.602)	0.176 (0.679)	0.696 (0.689)	0.31 (0.401)	0.99 (0.846)	0.299 (0.097 , 0.501)	0.004	0.497 (0.227 , 0.767)	<0.001	0.796 (0.526 , 1.066)	<0.001
VBSI	0.343 (0.863)	0.934 (0.925)	0.152 (1.461)	1.363 (1.397)	0.482 (1.178)	2.24 (1.889)	0.718 (0.418 , 1.076)	<0.05	1.327 (0.791 , 1.915)	<0.05	2.045 (1.408 , 2.649)	<0.05
GM Atrophy	-0.036 (0.296)	0.174 (0.329)	-0.004 (0.393)	0.311 (0.583)			0.245 (0.114 , 0.376)	<0.001	0.315 (0.117 , 0.512)	0.002	0.56 (0.356 , 0.764)	<0.001
WM Atrophy	1.107 (0.407)	1.436 (0.541)	1.016 (0.477)	1.735 (0.739)			0.341 (0.141 , 0.541)	0.001	0.626 (0.360 , 0.891)	<0.001	0.967 (0.619 , 1.315)	<0.001
Between-group differences in direct measures of change were calculated using generalised least squares regression, adjusted for age, gender, study site and scan interval and reported with bias-corrected 95% CIs. Grey highlights significant (p<0.05) between-group differences.												

Table 18-3. 6-, 9- and 15-month effect size estimates from the PADDINGTON study.			
	Effect Size Estimate (95% CI)		
	6 months	9 months	15 months
Cognitive Measures			
Letter Fluency	0.133 (-0.4, 0.603)	0.618 (-0.074, 1.183)	0.664 (-0.031, 1.32)
Category Fluency	0.226 (-0.212, 0.663)	0.125 (-0.417, 0.662)	0.348 (-0.204, 0.892)
HVLT delayed recall	0.491 (-0.006, 0.926)	0.001 (-0.525, 0.531)	0.502 (-0.122, 1.033)
HVLT total correct	0.115 (-0.357, 0.589)	0.119 (-0.329, 0.609)	0.208 (-0.181, 0.584)
HVLT Recognition	0.189 (-0.147, 0.449)	-0.264 (-0.692, 0.075)	-0.159 (-0.837, 0.32)
Symbol Digit Modalities	0.637 (0.08, 1.154)	0.343 (-0.108, 0.808)	0.799 (0.344, 1.254)
Trail A Time (seconds)	0.209 (-0.103, 0.471)	-0.055 (-0.369, 0.31)	0.212 (-0.115, 0.574)
Trail B Time (Seconds)	0.114 (-0.265, 0.443)	-0.23 (-0.678, 0.164)	-0.073 (-0.492, 0.25)
Stroop Word	0.285 (-0.086, 0.573)	0.063 (-0.257, 0.448)	0.305 (-0.081, 0.614)
Stroop Colour	0.25 (-0.191, 0.684)	0.192 (-0.227, 0.594)	0.36 (-0.026, 0.705)
Stroop Interference	0.176 (-0.194, 0.539)	0.303 (-0.11, 0.685)	0.492 (-0.026, 0.943)
Clinical Measures			
TMS (square root)	0.047 (-0.473, 0.609)	0.58 (0.087, 1.096)	0.545 (0.075, 1.123)
TFC score *	0.337 (-0.532, 1.325)	0.176 (-1.05, 1.323)	0.388 (-0.478, 1.241)
Micro-Structural Neuroimaging Measures			
Caudate FA	0.37 (-0.125, 0.826)	0.285 (-0.109, 0.651)	0.518 (0.12, 0.878)
Caudate MD (mm ² /s)	0.539 (0.199, 0.83)	0.62 (0.173, 1.029)	1.111 (0.77, 1.434)
Caudate RD (mm ² /s)	0.522 (0.178, 0.815)	0.611 (0.178, 1.017)	1.067 (0.729, 1.387)
Caudate AD (mm ² /s)	0.562 (0.211, 0.859)	0.627 (0.145, 1.059)	1.174 (0.839, 1.493)
Putamen FA	-0.036 (-0.359, 0.301)	-0.206 (-0.558, 0.152)	-0.269 (-0.654, 0.139)
Putamen MD (mm ² /s)	0.433 (0.147, 0.72)	0.286 (-0.074, 0.636)	0.723 (0.384, 1.017)
Putamen RD (mm ² /s)	0.328 (0.06, 0.603)	0.217 (-0.154, 0.562)	0.567 (0.231, 0.868)
Putamen AD (mm ² /s)	0.553 (0.229, 0.851)	0.379 (0.034, 0.723)	0.92 (0.526, 1.261)
WM FA	0.234 (-0.164, 0.638)	-0.089 (-0.477, 0.285)	0.165 (-0.284, 0.65)
WM MD (mm ² /s)	0.499 (0.072, 0.93)	0.188 (-0.148, 0.536)	0.615 (0.197, 1.095)
WM RD (mm ² /s)	0.39 (-0.053, 0.785)	0.103 (-0.21, 0.419)	0.506 (0.08, 0.944)
WM AD (mm ² /s)	0.495 (0.075, 0.887)	0.275 (-0.074, 0.813)	0.614 (0.212, 1.139)
CC FA	0.432 (0.112, 0.818)	0.148 (-0.206, 0.471)	0.681 (0.17, 1.147)
CC MD (mm ² /s)	0.249 (-0.183, 0.763)	0.147 (-0.304, 0.562)	0.304 (-0.117, 0.899)
CC RD (mm ² /s)	0.371 (-0.004, 0.876)	0.102 (-0.318, 0.498)	0.413 (-0.048, 1.034)
CC AD (mm ² /s)	0.015 (-0.351, 0.379)	0.235 (-0.235, 0.641)	0.211 (-0.157, 0.719)
Macro-Structural Neuroimaging Measures			
Caudate BSI (%Baseline)	0.697 (0.359, 1.021)	0.638 (0.324, 0.981)	1.191 (0.742, 1.687)
Whole-brain BSI (% Baseline)	0.479 (0.157, 0.774)	0.698 (0.314, 1.064)	0.865 (0.465, 1.199)
Putamen Vol (% Baseline)	0.101 (-0.187, 0.397)	0.538 (0.2, 0.899)	0.777 (0.331, 1.183)
CC Vol (% Baseline)	0.113 (-0.272, 0.557)	0.170 (-0.213, 0.609)	0.205 (-0.191, 0.63)
Ventricular BSI (mls)	0.791 (0.412, 1.143)	0.925 (0.553, 1.279)	1.025 (0.672, 1.323)
GM Atrophy (% Baseline)	0.772 (0.243, 1.23)	0.612 (0.304, 1.101)	0.86 (0.554, 1.219)
WM Atrophy (% Baseline)	0.615 (0.261, 1.028)	0.93 (0.566, 1.283)	0.957 (0.589, 1.325)
Cortical Thickness Measures			
Parietal Lobe Thickness (mm)	0.043 (-0.315, 0.415)	0.247 (-0.153, 0.647)	0.376 (-0.109, 0.855)
Occipital Lobe Thickness (mm)	0.295 (-0.161, 0.771)	0.22 (-0.196, 0.667)	0.512 (0.011, 0.997)
Estimate and 95% bias-corrected and accelerated CIs for effect sizes over 6, 9 and 15 months for differences between change in HD and controls. * TFC change relative to zero in HD group only.			

18.3.1 Short-Interval VBM

After stringent QC the 6-month VBM analysis included 74 participants (31 controls and 43 HD participants) and the 9-month analysis included 72 participants (30 controls and 42 HD participants). Significant atrophy in the early HD group compared with healthy controls was detectable in the WM over just 6- and 9-month intervals (Figure 18-2). Over six months this was limited to a very small region of the para-striatal WM. Over 9 months significant atrophy was detectable in the splenium, para-striatal and occipital WM. No significant between-group differences were detectable using this technique in the GM over these intervals.

WM SPMs

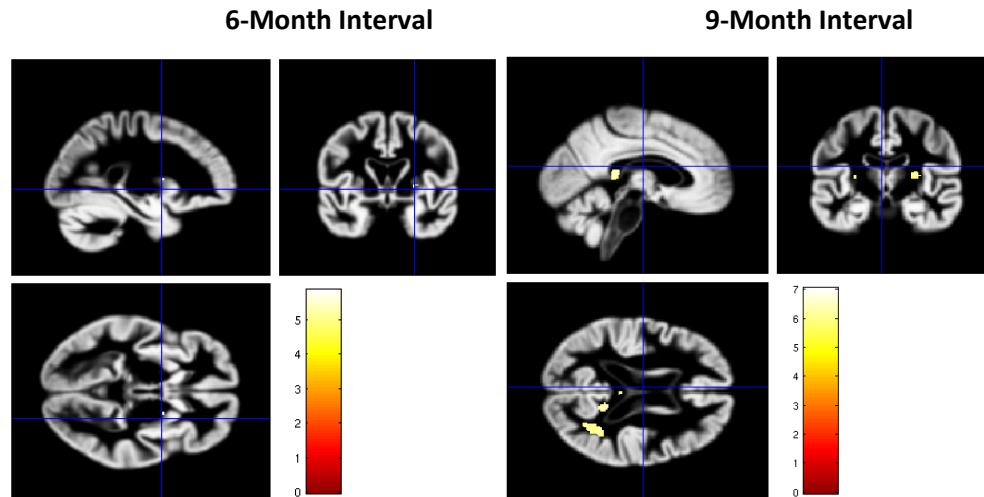


Figure 18-2. Regions of significant 6- (left) and 9-month (right) WM atrophy in the HD group compared with controls, adjusted for multiple comparisons using FWE at $p < 0.05$. The colour bars represent the T-scores.

18.4 Discussion

Cognitive, clinical and neuroimaging measures have been shown to be sensitive to pathology within neurodegenerative diseases such as HD, and are proposed as biomarker candidates. In this multi-site study of stage I HD the relative sensitivities of a battery of biomarker candidates were compared. The cross-sectional analysis suggests that the ROI examined may be more important than the imaging modality applied, with the subcortical GM regions outperforming the global measures and CC for both macro- and micro-structural metrics. The longitudinal results show for the first time that several measures are sensitive to disease progression in this cohort over as short an interval as six months, demonstrating the utility of these biomarkers as short-interval read-outs in clinical trials.

TMS and TFC are attractive as biomarkers in HD due to their direct relationship to the clinical manifestation of the disease. TFC however was not sensitive to change within this early-stage HD group and TMS required longer intervals than other measures to detect significant between-group differences and is therefore not ideal as a short interval outcome measure. Between-group differences in the cognitive measure were a result of a combination of decline in HD group performance but also an increase (practice effects) in the control group. These effect sizes therefore will be influenced by the number of visits and the previous experience in the cohort of the tests being applied.

Macro-structural neuroimaging measures provided the strongest longitudinal effect sizes of the battery tested; most notably the CBSI, VBSI and WM atrophy. The cortical thickness measures however were not sensitive to HD-related change, with only occipital lobe atrophy over 15 months reaching statistical

significance. Diffusion neuroimaging metrics were relatively sensitive, but effect sizes were on average not as high as those of the macro-structural measures. This is perhaps surprising; neurodegeneration in HD is a slow process where neurons undergo prolonged alterations including axonal- and dendritic-remodelling prior to gross morphometric change. Consequently it was hypothesized that the micro-structural diffusion metrics would show significant advantages in sensitivity over the macro-structural volumetrics. It may be that since atrophic changes are well-established by early clinical HD, the theoretical advantages of micro-structural metrics only play out when investigating pathology during the pre-manifest stages of the disease. Alternatively this may be a consequence of less developed diffusion image analysis tools in comparison to the volumetric biomarkers now available. Furthermore, from a technical perspective, the reliance of diffusion MRI on echo planar imaging confers on it a reduced signal-to-noise ratio and concomitant decrease in spatial resolution compared with T1-weighted imaging. These factors could potentially mitigate the advantages that the technique offers in detecting micro-structural abnormalities.

Diffusivity metrics (MD, AD and RD) out-performed FA within the GM of the caudate and putamen but this effect was inverted within the WM of the CC. The differential efficacy of the diffusivity versus anisotropy measures here is most likely due to the tissue characteristics of GM and WM. This finding recommends the use, longitudinally, of diffusivity measures for GM and FA for single WM tract analysis – the sensitivity of FA within the whole-brain WM was reduced, potentially due to a more complex WM tract structure.

In contrast to ROI-based volumetric biomarkers, VBM maps allow for detection and localisation of atrophy across the whole-brain. Significant WM atrophy was detected using VBM over six and nine months in regions consistent with previous findings over longer scan-intervals (Tabrizi et al. 2011); notably within and around the basal ganglia, splenium and occipital WM. It is suggested that WM VBM may provide valuable supplementary information to hypothesis-driven ROI-based biomarkers. For example, the detection of significant atrophy over just six months within the splenium is a finding not currently covered by any widely used ROI volumetric biomarker.

18.5 Guidelines for Clinical Trial Biomarker Selection in Early-Stage HD

The PADDINGTON study effect sizes have been used to develop guidelines (illustrated in Figure 18-3) to facilitate the choice of biomarkers for usage in future clinical trials of potentially disease-modifying therapies in early-stage HD. The use of these guidelines should be tailored to the hypothesized mechanism of action of the drug or therapy on trial.

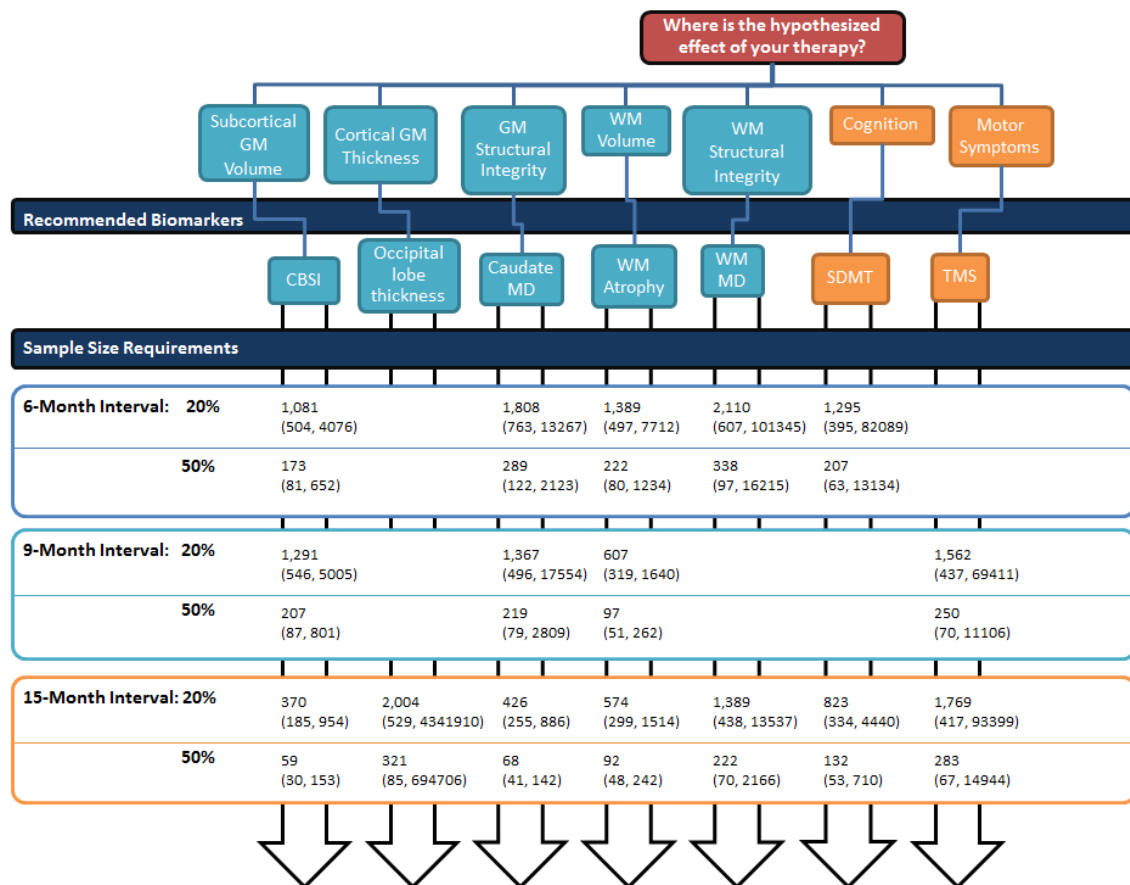


Figure 18-3. Guidelines for biomarker selection over short and varying time intervals in future clinical trials of potentially disease-modifying therapies for early stage HD. Sample size requirements are per treatment arm; calculated using the standard formula (Julious 2009), with 90% power and two-tailed $p < 0.05$, for therapies with 20% and 50% estimated treatment efficacy in early-stage HD.

Based on the superior sensitivity of several macro-structural neuroimaging measures over short time intervals their use is recommended as a means of obtaining early, informative read-outs in proof-of-concept studies over short scan intervals. Results here would facilitate the decision as to whether to continue the trial over longer intervals and potentially increase participant numbers. Once sufficiently powered, disease-modification could be demonstrated over longer time intervals in large scale efficacy studies using approved measures such as TMS as the primary end-point and specific neuroimaging metrics

as secondary end-points. An adaptive approach such as this based on early, meaningful data may improve the viability of disease-modifying clinical trials.

Utilisation of a range of markers is recommended to assess different aspects of treatment efficacy. TRACK-HD has provided strong evidence for the functional relevance of neuroimaging biomarkers, with associations shown between structural brain atrophy, genetic burden, TFC and TMS decline (Tabrizi et al. 2012). Attenuation of disease progression as measured by neuroimaging biomarkers may not however always be accompanied by functional improvement and vice versa. As surrogate biomarkers neuroimaging measures are only recommended for use alongside a direct clinical or cognitive biomarker.

In the context of biomarker candidates, when deciding between metrics and considering study design, other factors may come into play. Estimated study drop-out should be factored into sample size calculations in order to ensure that a sufficient level of power is maintained. Neuroimaging biomarkers are logistically complicated and expensive compared with standard clinical or cognitive end-points and their sensitivity is generally reduced in individuals with later-stage HD due to movement. Costs, expertise, equipment, time and ease of acquisition are therefore all essential considerations. It should also be decided whether a metric provides independent information on neurodegeneration. For example, in the current study, whole-brain (BBSI) and GM atrophy metrics performed very similarly in terms of effect sizes, which is not surprising since these structures include a substantial amount of the same information; hence there may be limited advantages in including measures from both regions in a study or trial. Conversely, although the cortical thickness metrics produced smaller effect sizes, these measures may provide unique information on neurodegeneration not detected by other (subcortical or diffusivity) biomarkers.

It is important to acknowledge the potential limitations of this approach. Although the PADDINGTON study was designed to imitate the set-up of a clinical trial there are unavoidable differences which may affect the accuracy of the effect size estimations given: the between-group comparison in clinical trials will be drug versus placebo rather than HD versus controls; a therapy might not just slow progression but might improve function; clinical trials will require more sites therefore inter-site consistency and reliability will become more of a factor; there will also be more assessments – most likely increasing practice effects in the cognitive measures most significantly. Additionally, non-HD related confounds might be present in observational neuroimaging studies. There is some suggestion that dehydration, hydrocephalus and diuretic therapy may affect regional volumetrics, particularly the ventricles (Schott et al. 2005). Other medications may also confound disease-related brain changes. None of the participants in the current study were enrolled in clinical trials, although many were on medications which target the central nervous system (CNS). This study was not designed to examine the specific effects of medication on each outcome; however, medication usage is acknowledged as a potential confounder. Nevertheless, mean dosages of

CNS-targeting drugs were relatively low, with overlap in usage between groups. Changes in medication over the study intervals were minimal for both groups, providing confidence that our findings are driven by HD-related pathology.

In a clinical trial of a pharmaceutical therapeutic the likelihood of drug-induced non-disease-modifying brain alterations is increased, for example from oedema, inflammation, increased cytoplasmic volume or cell membrane thickness (Gilman et al. 2005). Additionally, injection of a therapeutic into the brain could physically alter brain morphometry. Under any of these circumstances volumetric neuroimaging biomarkers would be invalidated as a means to track pathological progression, although might have utility as a safety measure. Furthermore, neuroimaging read-outs may not be suitable for all types of intervention; their value will be dependent on the mechanism of action of the therapy, together with the time required for it to mediate an effect.

Nevertheless, neuroimaging measures have been shown to be the strongest biomarker candidates for detecting hypothetical treatment effects over six months. Hence they may provide valuable short-term read-outs in future clinical trials in HD. It is hoped that careful study design based on estimates of sample size requirements from either TRACK-HD, PREDICT-HD or the PADDINGTON study, will facilitate highly powered, cost-efficient clinical trials in HD in the near future.

19 PART 3: A Summary

A comprehensive comparison of the performance of clinical, cognitive and neuroimaging biomarkers in early-stage HD over short time intervals was conducted as part of the PADDINGTON study and reported in Part 3 of this thesis. Effect sizes were used to provide estimates of sample size requirements for clinical trials utilising one or more of these measures as outcome variables. This study supplements the effect and sample size estimates published from the TRACK-HD and PREDICT-HD cohorts.

Macro-structural, and to a lesser extent micro-structural, biomarkers were shown to be sensitive to HD-related change over clinically relevant short time intervals. Neuroimaging metrics are however surrogate markers of disease progression and therefore only appropriate as secondary outcome measures. Direct clinical measures of disease progression, such as the TMS, would be required as the primary end-point. Neuroimaging biomarkers are therefore recommended as a means of obtaining early, confidence-instilling read-outs over short scan intervals, facilitating the decision as to whether to continue the trial over longer intervals and possibly also increase patient numbers. Informing this decision is particularly valuable due to the high costs of conducting these trials. Once sufficiently powered, disease-modification could be demonstrated over longer time intervals using approved, direct clinical measures such as TMS.

20 Conclusions

These are times of great optimism in HD research. In the near future it is hoped that years of work on therapeutic and biomarker development will be realised in well-designed, highly powered trials of potentially disease-modifying therapies. Whether or not the inclusion of neuroimaging biomarkers is warranted, clear and robust guidelines for their use are available from TRACK-HD, PREDICT-HD and now the PADDINGTON study. As well as reporting the results of the PADDINGTON study biomarker comparison this thesis has also evaluated automated tools sensitive to neurodegenerative change, which may facilitate large-scale clinical trial volumetric analysis, and enhanced current knowledge of the HD phenotype and its underlying neuropathology.

Large cohorts will be required to sufficiently power clinical trials in HD to detect treatment efficacy. Therefore there is a need for identification of robust fully-automated methods, comparable to the current semi-automated or manual gold-standards, which would facilitate large-scale volumetric analysis. Several automated methods are emerging as viable alternatives or supplementary measures to these gold-standards. Fully automated techniques are often thought of as being unbiased, objective and reproducible by different operators but this has been shown not to be the case with VBM (Henley et al. 2010) and FreeSurfer analysis (Section 9.4) as varying the parameters can significantly affect the results produced. Consequently it is very important that parameters are justified and clearly stated in all reports. Visual inspection of outputs is also deemed necessary. Manual edits to these segmentations may be a reasonable compromise between the capacity to analyse large datasets but also maintaining an acceptable level of quality. Overall these results highlight the fact that a brain region is only a good biomarker target if it can be reliably measured and this is additionally dependent on the measurement technique applied. Every technique requires validation on the cohort and imaging data under investigation.

Studies within this thesis have led to several advancements in our understanding of HD, adding to our overall knowledge of the brain and strengthening the argument for the clinical relevance of neuroimaging measures as surrogate end-points in HD. The novel findings reported include: firstly, that cerebellar GM and WM pathology may have a potential role in the motor and psychiatric symptoms of HD; secondly, the HD emotion recognition deficit varies significantly between stimulus modalities (photos, vocal expressions and film clips) and the severity of this impairment associates with widespread WM pathology; and lastly, that there is a significant association between occipital cortical thickness and cognitive task performance. To have the ability to understand and interpret clinical and neurological pathology will facilitate future research and good clinical practice.

Effect sizes and sample size requirements were reported for a comprehensive battery of clinical, cognitive and (macro- and micro-structural) neuroimaging biomarkers, which clearly demonstrated the power of

neuroimaging biomarkers over short time intervals. Guidelines are provided for selection of appropriate, focused biomarkers over specified periods with varying hypothesized therapeutic efficacies. These guidelines supplement the previously published TRACK-HD and PREDICT-HD sample size estimations, as the PADDINGTON study assessed all biomarkers over shorter, more clinically relevant, time intervals in early-stage HD; the most likely cohort to be recruited for the next clinical trials.

In conclusion, neuroimaging biomarkers, particularly macro-structural volumetric measures, are unrivalled in their sensitivity to HD-related change over short time intervals in early-stage HD. These are therefore strong biomarker candidates for future clinical trials in HD. As imaging metrics are an indirect measure of neuronal activity/health, which cannot be assumed to relate directly to clinical progression, other measures will be needed alongside them, e.g. TMS, TFC and/or quality-of-life scales. The advantages and disadvantages of the inclusion of imaging in clinical trials must be seriously considered; thorough study design and set-up with careful biomarker selection are imperative to ensure the appropriate treatment of sensitive clinical trial data. Evidence-based guidelines and recommendations for biomarker use in HD are now available and the research field is ready for large-scale clinical trials of potentially disease-modifying therapies; and one step closer to finding a cure for this debilitating disease.

21 Publications

Published peer-reviewed papers based on results described in this thesis are referenced below. I am grateful to all individuals involved in these publications, whose contributions are listed below each publication.

CHAPTER 2

Rees EM, Scahill RI & Hobbs, NZ. Longitudinal Neuroimaging in Huntington's Disease. *Journal of Huntington's Disease* 2013;2:21–39.

Drafting and critical revision of the manuscript: ER, NH, RIS

CHAPTER 10

Rees EM, Farmer R, Cole JH, Henley SM, Sprengelmeyer R, Frost C, Scahill RI, Hobbs NZ & Tabrizi SJ. Inconsistent Emotion Recognition Deficits across Stimulus Modalities in Huntington's Disease. *Neuropsychologia* 2014; 64C:99-104.

For this publication I recruited participants from the London site of the PADDINGTON study. My concept and design of the novel cognitive test was facilitated by SMH. I conducted the imaging data analysis with advice from JC, NH and RIS. RF conducted all the statistical analyses and contributed a significant amount to the interpretation of the data.

Study concept and design: ER, RF, NH, SMH, RIS

Acquisition of data: ER, NH, RS

Analysis and interpretation of data: ER, NH, RIS, JC, RF

Drafting of the manuscript: ER, RF, NH

Critical revision of the manuscript for important intellectual content: ER, RF, JC, SMH, RS, CF, RIS, NH, ST

Statistical analysis: RF, CF

CHAPTER 13

Rees EM, Farmer R, Cole JH, Haider S, Durr A, Landwehrmeyer B, Scahill RI, Tabrizi SJ & Hobbs NZ. Cerebellar Abnormalities in Huntington's Disease: A Role in Motor and Psychiatric Impairment? *Movement Disorders* 2014; doi: 10.1002/mds.25984.

For this study I led on the study concept and design, with guidance from NH. I developed a novel method for volumetric analysis of the cerebellum with support from NH and RIS, conducted the diffusion analysis with advice from JC and ran the statistical analyses with supervision from RF.

Study concept and design: ER, NH, ST, BL

Acquisition of data: NH, ER, JC, SH, AD

Analysis and interpretation of data: ER, NH, RF, RS, JC

Drafting of the manuscript: ER, NH, RF, RS

Critical revision of the manuscript for important intellectual content: ER, NH, RF, RS, ST, SH, JC, AD, BL

Statistical analysis: RF, ER

CHAPTER 18

Hobbs NZ, Cole JH, Farmer RE, Rees EM, Crawford HE, Malone IB, Roos RAC, Sprengelmeyer R, Durr A, Landwehrmeyer B, Scahill RI, Tabrizi SJ & Frost C. Evaluation of multi-modal, multi-site neuroimaging measures in Huntington's disease: Baseline results from the PADDINGTON study. *Neuroimage: Clinical* 2013;2:204-11.

This was a large multi-site study involving a large number of PADDINGTON study investigators. I was responsible for recruiting participants for the London site, data collection during study visits, volumetric image analysis, interpretation of data and critical review of the manuscript.

Obtained funding: SJT, BL

Study concept and design: NH, SJT, BL, CF

Acquisition of data: NH, ER, JC

Analysis and interpretation of data: NH, JC, RF, ER, RS, SJT, CF

Statistical analysis: RF, CF

Drafting of the manuscript: NH

Critical revision of the manuscript for important intellectual content: NH, JC, RF, ER, HC, IM, RR, RS, AD, BL, RS, SJT, CF

Hobbs NZ, Farmer RE, Rees EM, Cole JH, Haider S, Malone IB, Sprengelmeyer R, Johnson H, Mueller H-P, Sussmuth SD, Roos RAC, Durr A, Frost C, Scahill RI, Landwehrmeyer B & Tabrizi SJ. Variable short-interval observational data to inform clinical trial design in early Huntington's Disease. – under review at *the Journal of Neurology, Neurosurgery & Psychiatry*.

This is the longitudinal follow-up to the baseline PADDINGTON publication and therefore, again, involved a large number of study investigators. I was responsible for data collection during study visits, volumetric image analysis, interpretation of data and critical review of the manuscript.

Study concept and design: NH, CF, RIS, BL, SJT

Acquisition of data: NH, JHC, ER, SH, IBM, RS, HJ, HPM, SDS, RACR, AD, RIS

Analysis and interpretation of data: NH, RF, CF, SJT, ER, JHC, SJT

Statistical Analysis: RF, CF

Drafting of the manuscript: NH

Critical revision of the manuscript for important intellectual content: NH, RF, JHC, ER, SH, IBM, RS, HJ, HPM, SDS, RACR, AD, CF, RIS, BL, SJT

22 Acknowledgements

I am grateful to all of the participants who generously gave their time to be part of this research. I would also like to thank the European Union for funding the PADDINGTON study (contract n. HEALTH-F2-2010-261358), CHDI for funding the TRACK-HD study and the Department of Health's NIHR Biomedical Research Centres (BRC) funding scheme which supports research at UCL.

Thank you to all the PADDINGTON and TRACK-HD staff. In particular I would like to thank Rachael for being such a supportive and encouraging supervisor and for all your guidance with writing this thesis; Nicky for your incredibly valuable ideas, suggestions and critique; Sarah for giving me the opportunity to work in your group on such a wide range of interesting projects; Helen and Eli for your help with image analysis and QC; James for your advice and assistance with the diffusion analysis; and a huge thank you to Ruth for your statistical input and patience.

Finally, I want to thank my parents for always encouraging me to do what I enjoy, and Suraj for your patience and pep talks during the completion of this thesis.

23 Appendix

23.1 UHDRS: TMS Scoring System

A TMS ranges between 0 and 124. This score is made up of ratings from the following subscales:

- Ocular Pursuit (horizontal)
 - 0-complete
 - 1-jerky
 - 2-interrupted/full range
 - 3-incomplete range
 - 4-cannot pursue
- Ocular Pursuit (vertical)
 - 0-complete
 - 1-jerky
 - 2-interrupted/full range
 - 3-incomplete range
 - 4-cannot pursue
- Saccade Initiation (horizontal)
 - 0-normal
 - 1-increased latency
 - 2-suppressible blinks/head movements to initiate
 - 3-unsuppressible head movements
 - 4-cannot initiate
- Saccade Initiation (vertical)
 - 0-normal
 - 1-increased latency
 - 2-suppressible blinks/head movements to initiate
 - 3-unsuppressible head movements
 - 4-cannot initiate
- Saccade Velocity (horizontal)
 - 0-normal
 - 1-mild slowing
 - 2-moderate slowing
 - 3-severely slow, full range
 - 4-incomplete range
- Saccade Velocity (vertical)
 - 0-normal
 - 1-mild slowing
 - 2-moderate slowing
 - 3-severely slow, full range
 - 4-incomplete range
- Dysarthria
 - 0-normal
 - 1-unclear, no need to repeat
 - 2-must repeat
 - 3-mostly incomprehensible
 - 4-mute
- Tongue Protrusion
 - 0-normal
 - 1-<10 seconds

- 2-<5 seconds
- 3-cannot fully protrude
- 4-cannot beyond lips
- Finger Taps (right)
 - 0-normal (15/5sec)
 - 1-mild slowing or reduction in amp.
 - 2-moderately impaired. may have occasional arrests (7- 10/15sec)
 - 3-severely impaired. Frequent hesitations and arrests
 - 4-can barely perform
- Finger Taps (left)
 - 0-normal (15/5sec)
 - 1-mild slowing or reduction in amp.
 - 2-moderately impaired. may have occasional arrests (7- 10/15sec)
 - 3-severely impaired. Frequent hesitations and arrests
 - 4-can barely perform
- Pronate/Supinate (right)
 - 0-normal
 - 1-mild slowing/irregular
 - 2-moderate slowing and irregular
 - 3-severe slowing and irregular
 - 4-cannot perform
- Pronate/Supinate (left)
 - 0-normal
 - 1-mild slowing/irregular
 - 2-moderate slowing and irregular
 - 3-severe slowing and irregular
 - 4-cannot perform
- Fist-Hand-Palm Sequence
 - 0->4 in 10 seconds without cues
 - 1-<4 in 10 sec. without cues
 - 2->4 in 10 sec. with cues
 - 3-<4 in 10 sec. with cues
 - 4-cannot perform
- Rigidity-arms (right)
 - 0-absent
 - 1-slight or only with activation
 - 2-mild/moderate
 - 3-severe, full range of motion
 - 4-severe with limited range
- Rigidity-arms (left)
 - 0-absent
 - 1-slight or only with activation
 - 2-mild/moderate
 - 3-severe, full range of motion
 - 4-severe with limited range
- Bradykinesia
 - 0-normal
 - 1-minimally slow
 - 2-mildly but clearly slow
 - 3-moderately slow
 - 4-marked slowing, long delays in initiation
- Maximal Dystonia (trunk)

0-absent
 1-slight/intermittent
 2-mild/common or moderate/intermittent
 3-moderate/common
 4-marked/prolonged
 Maximal Dystonia (right upper extremity)
 0-absent
 1-slight/intermittent
 2-mild/common or moderate/intermittent
 3-moderate/common
 4-marked/prolonged
 Maximal Dystonia (left upper extremity)
 0-absent
 1-slight/intermittent
 2-mild/common or moderate/intermittent
 3-moderate/common
 4-marked/prolonged
 Maximal Dystonia (right lower extremity)
 0-absent
 1-slight/intermittent
 2-mild/common or moderate/intermittent
 3-moderate/common
 4-marked/prolonged
 Maximal Dystonia (left lower extremity)
 0-absent
 1-slight/intermittent
 2-mild/common or moderate/intermittent
 3-moderate/common
 4-marked/prolonged
 Maximal Chorea (Face)
 0-absent
 1-slight/intermittent
 2-mild/common or moderate/intermittent
 3-moderate/common
 4-marked/prolonged
 Maximal Chorea (buccal-oral-lingual)
 0-absent
 1-slight/intermittent
 2-mild/common or moderate/intermittent
 3-moderate/common
 4-marked/prolonged
 Maximal Chorea (Trunk)
 0-absent
 1-slight/intermittent
 2-mild/common or moderate/intermittent
 3-moderate/common
 4-marked/prolonged
 Maximal Chorea (right upper extremity)
 0-absent
 1-slight/intermittent
 2-mild/common or moderate/intermittent
 3-moderate/common

- 4-marked/prolonged
- Maximal Chorea (left upper extremity)
 - 0-absent
 - 1-slight/intermittent
 - 2-mild/common or moderate/intermittent
 - 3-moderate/common
 - 4-marked/prolonged
- Maximal Chorea (left lower extremity)
 - 0-absent
 - 1-slight/intermittent
 - 2-mild/common or moderate/intermittent
 - 3-moderate/common
 - 4-marked/prolonged
- Maximal Chorea (right lower extremity)
 - 0-absent
 - 1-slight/intermittent
 - 2-mild/common or moderate/intermittent
 - 3-moderate/common
 - 4-marked/prolonged
- Gait
 - 0-normal narrow base
 - 1-wide base, and/or slow
 - 2-wide base, walks with difficulty
 - 3-walks with assistance
 - 4-cannot attempt
- Tandem Walking
 - 0-normal for 10 steps
 - 1-1-3 deviations
 - 2->3 deviations
 - 3-cannot complete
 - 4-cannot attempt
- Diagnostic Confidence Level:
 - 0-normal (no abnormalities)
 - 1-non-specific motor abnormalities (less than 50% confidence)
 - 2-motor abnormalities that may be signs of HD (50-89% confidence)
 - 3-motor abnormalities that are likely signs of HD (90-98% confidence)
 - 4-motor abnormalities that are unequivocal signs of HD ($\geq 99\%$ confidence).

23.2 UHDRS: TFC Scoring System

Functional capacity is scored out of 13 based in ratings on the following subscores:

- Occupation
 - 0-unable
 - 1-marginal work only
 - 2-reduced capacity for usual job
 - 3-normal
- Finances
 - 0-unable
 - 1-major assistance
 - 2-slight assistance
 - 3-normal
- Domestic Chores
 - 0-unable
 - 1-impaired
 - 2-normal
- Activities of Daily Living*
 - 0-total care
 - 1-gross tasks only
 - 2-minimal impairment
 - 3-normal
- Care Level
 - 0-full time skilled nursing
 - 1-home or chronic care
 - 2-home

* Activities of daily living include daily self-care activities such as feeding ourselves, bathing, dressing, grooming, work and leisure.

23.3 TRACK-HD Acquisition Parameters

In the TRACK-HD study 3D MPRAGE T1-weighted image acquisition sequences were used on 3T Siemens and Philips whole-body imagers with the following imaging parameters: TR = 2200ms (Siemens)/ 7.7ms (Philips), TE=2.2ms (S)/3.5ms (P) , FA=10 (S)/8 (P), FOV =28cm (S)/ 24cm (P), matrix size 256x256(S)/224x224(P), 208(S)/164(P) sagittal slices to cover the entire brain with a slice thickness of 1.0 mm with no gap.

23.4 Manual Putamen Delineation: Standard Operational Procedure (SOP)

Section 7.2 describes the development of a manual delineation protocol for volumetric analysis of the putamen. The full protocol is detailed below. Prior to putamen analysis the whole-brain region for the corresponding scan must have been segmented and the scan registered to standard MNI305 space.

1. Start MIDAS software with the command: `wp2-midas -morph 90 112`. This automatically calculates 90% and 112% of the mean brain intensity, required for the thresholds.
2. Load up a standard-space registered brain region by clicking on *Regions>ShowDatabase>Register-Template-9dof6* in the top menu. Search for the brain region you want by typing the participant number or a # followed by the scan number in the search box at the top.
3. Highlight the brain region by clicking within the outline and select *Measure->Simple Mean*. This calculates the mean voxel intensity over the whole-brain region. Values can be seen in the *Log* window.
4. Remove the brain region; *Region>Remove*.
5. Select *Region>Edit>Irregular Volume>2 views*.
6. Move to axial view in the main window and zoom x9. In the editor window select the sagittal view and zoom x4.
7. Set the upper and lower thresholds to 112% and 90% of the brain mean intensity (found in *Log* window).
8. Move to a slice in which the putamen appears clearly.
9. Seed both left and right structures. Sometimes several seeds are required on one slice, especially if the intensity of the putamen is inhomogeneous, e.g. Figure 23-1a. Make sure to seed the posterior 'tail'.
10. Use the 'Draw' tool to edit the borders, e.g.
 - Run down the length of the WM between the putamen and claustrum (example in Figure 23-1b).
 - At this point ignore any gaps and simply concentrate on a rough border outline.
 - Any "projections" or tissue "bridges" between the caudate and putamen should not be included.

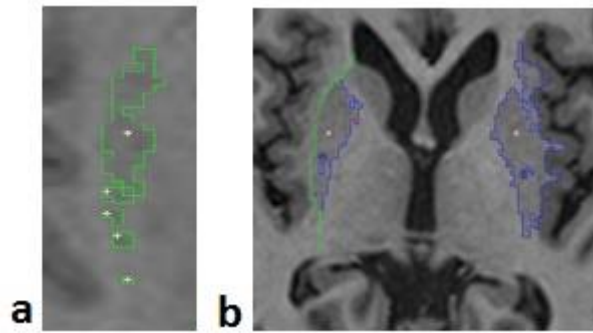


Figure 23-1. a) Several seeds may be needed to roughly fill the structure on each slice; b) Manually edit the lateral border, disconnecting the putamen from the WM and claustrum.

11. Move up the slices continuing the segmentation. Keep going until you reach the top, using the position tool and the editor (sagittal) view to judge where the top is.
12. Scroll inferiorly and continue the segmentation down to the last slice in which the inferior capsule clearly separates the putamen from the caudate. Examples of this are shown in Figure 23-2. This may be different levels in the left and right hemispheres.

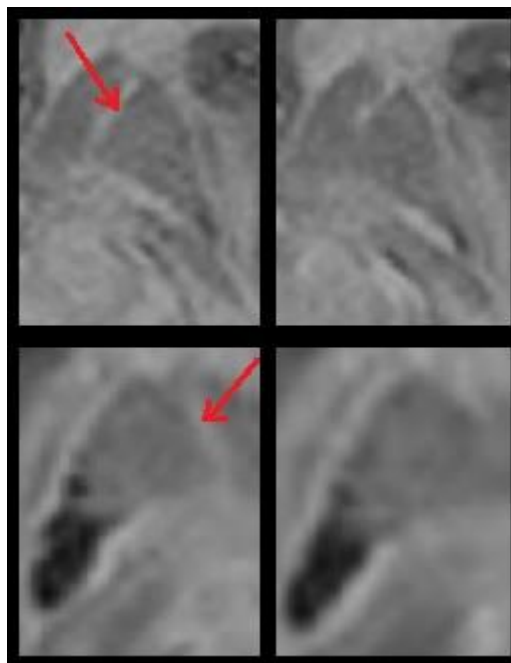


Figure 23-2. Two examples of the most inferior slice in which the putamen should be segmented (left) – here the putamen is still clearly separated by WM from the caudate (red arrows). Once this separation is no longer clear (e.g. right images) stop the segmentation and do not segment this slice.

13. Select 'App. Thresh.'.
14. Go back up through the segmentation tidying each slice:
 - Smooth edges and fill gaps that look biological implausible.

- There are often dark spots/lesions within the putamen. These should be included if they are within the body of the putamen or are continuous with its border.
- If the tail and head of the putamen are connected (not disconnected by WM or black vessels) segment the whole structure. If a disconnect is apparent only segment the head (Figure 23-3).

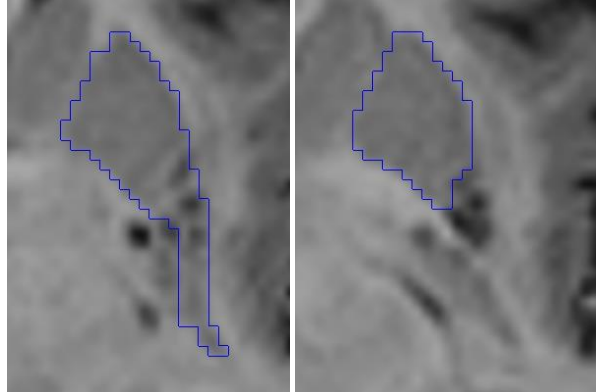


Figure 23-3. If the tail and head of the putamen are connected on that slice segment the whole structure (left). When this becomes split by WM deseed the tail (right).

15. Stay with the axial view in the main window. Select the Position tool and find the most medial slice including right putamen.
16. Using both the axial and sagittal views do more detailed edits:
 - Ensure that both views have smooth biologically plausible outlines (e.g. Figure 23-4).

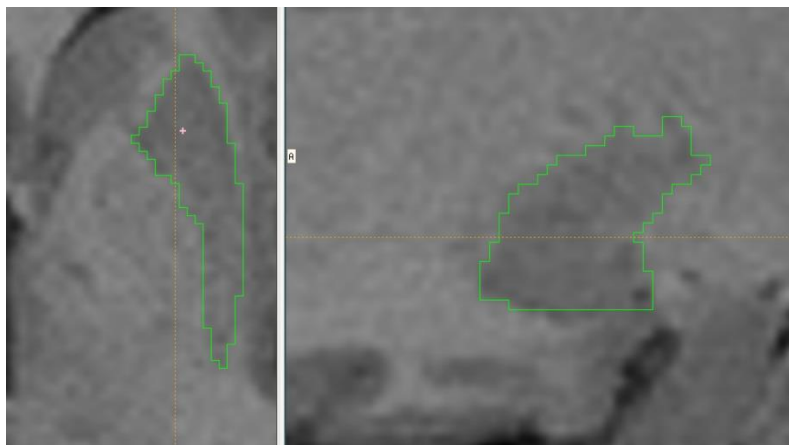


Figure 23-4. Ensure that edges are smooth and biologically plausible in both views.

- Continue laterally editing each slice.
 - Do the same for the left putamen.
17. Save the segmentation; click OK in the editor window, in the main window highlight the region, *Regions>Database In>e.g. Putamen.*

23.5 Manual Cerebellum Delineation: SOPs

Section 8.3 describes the development of a manual delineation protocol for volumetric analysis of the cerebellum. The full protocol is detailed below. It should be noted that, due to inter-scanner differences in voxel intensity and tissue contrast, it is recommended to optimise the lower intensity threshold applied in the following protocol based on the data under investigation. Prior to cerebellar analysis the whole-brain region for the corresponding scan must have been segmented and the scan registered to standard MNI305 space.

1. Open a terminal window and type: *hdni-midas -morph <70> 160*
2. Load up the standard-space registered brain region from *Regions>Show Database>Register-Template-9dof6*.
3. Remove the whole-brain region by highlighting it in red and selecting *Regions>Remove*.
4. Open the Irregular Volume Segmentation tool by clicking on *Regions>Edit>Irregular Volume>2 views*.
5. In the new window, tick the 'threshold' box.
6. Set the upper threshold at the 160% of the mean brain intensity (value displayed in the Log window).
7. Set the lower threshold value at the empirically established optimal level. In the data tested in Section 8.3.1 of this thesis this was found to be 65% for Philips scans and 70% for Siemens scans.
8. Select the coronal view in the main Midas window and the sagittal view in the Irregular Volume window.

The two protocol options deviate from this point onwards:

23.5.1 Protocol 1

9. Find the starting point: scroll towards the anterior of the cerebellum stopping on a slice which shows definite cerebellar tissue (Figure 23-5A).

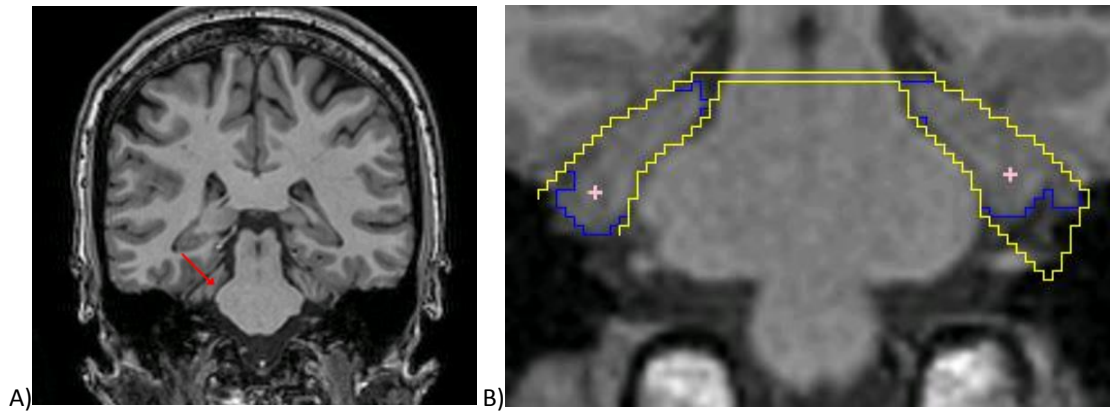


Figure 23-5. A) An example of a starting slice. Seed the superior anterior regions of cerebellum. B) Cut around the top edge of the cerebellum removing the cerebrum and middle to remove all brainstem.

10. Add a seed to both sides of the cerebellum.
11. With the 'Poly' tool cut around the top edge removing the cerebrum, and middle to remove all brainstem (Figure 23-5B).
12. Take care not to include the cranial nerve that runs between the cerebellum and cerebrum (red arrows in Figure 23-6).

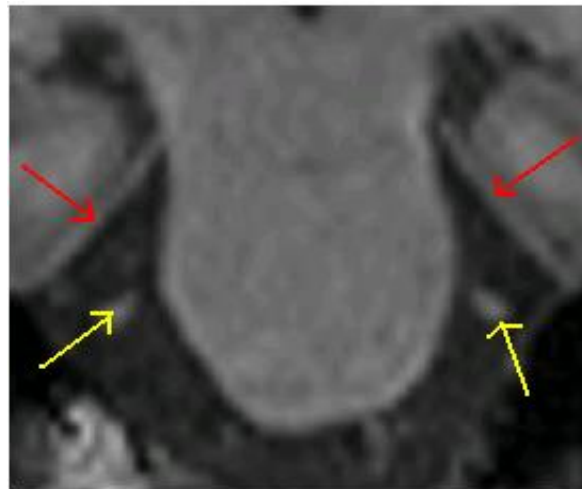


Figure 23-6. Highlights nerves that may be confused for cerebellar GM but should not be included in the segmentations.

13. Seed the inferior cerebellum as it emerges.
 - Do not confuse nerve for cerebellum (yellow arrows in Figure 23-6). These nerves are paler than cerebellar GM.
 - If it is unclear as to whether it is GM or nerve, check in the sagittal view and include/exclude depending on which looks better.

- If GM and nerve are being included together by the intensity thresholds, manually 'Draw' around the nerve to remove it.
14. Scroll anteriorly, repeating the manual removal of the cerebrum and brainstem, until the most anterior point of the cerebellum is reached.
 15. Scroll posteriorly, continuing the manual segmentation removing cerebrum, brainstem and any unwanted areas at the sides/inferior borders. The connective vermis will appear at the top between the hemispheres. Be sure to add a seed to this (Figure 23-7).

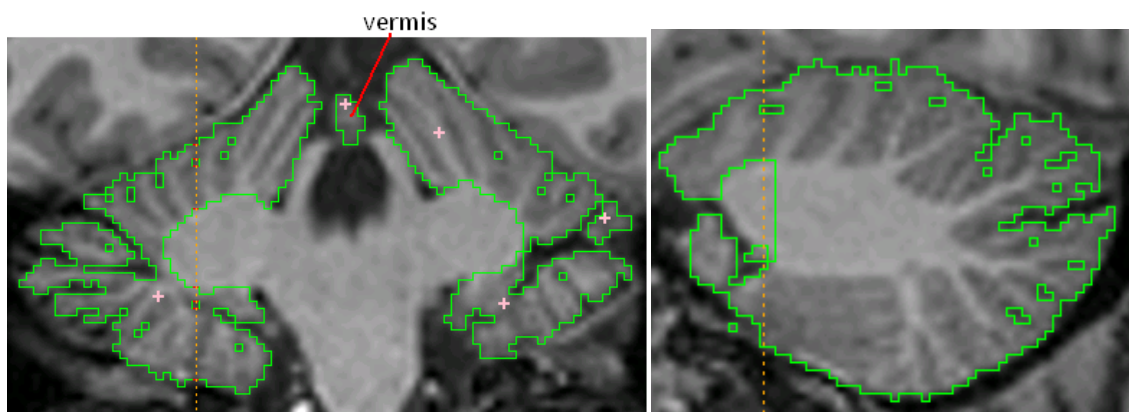


Figure 23-7. An example of WM removal before the hemispheres join. Note the seeded vermis between the hemispheres. The same segmentation is shown in the coronal and sagittal views.

- Sometimes seeds do not take therefore it is necessary to draw a rough outline along the bottom of the cerebellum.
16. When the superior cerebellar GM joins, this protocol changes (demonstrated in Figure 23-8).

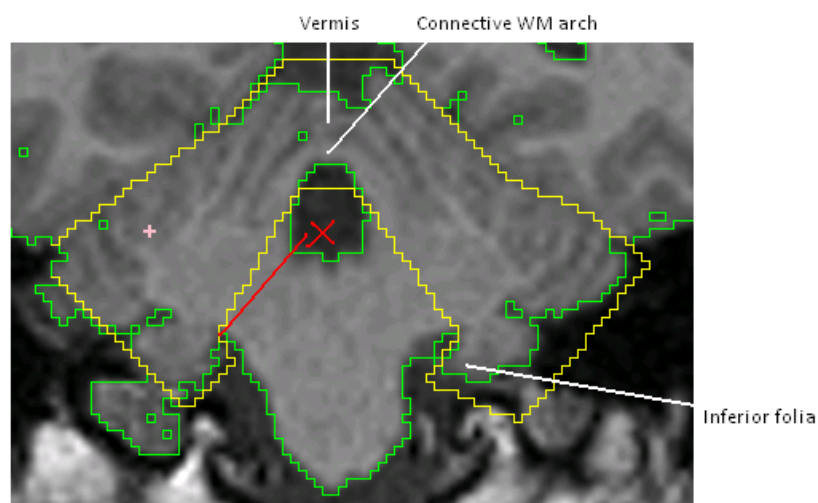


Figure 23-8. When the GM of the hemispheres become connected by the vermis, include this and the underlying WM arch. Continue the curve of the connective WM down to the inferior folia (yellow line is correct, the red segmentation line is wrong).

- The connecting vermis must be included.
- At this point include the connective WM beneath the vermis and that within the cerebellar hemispheres. The outline should be made, again with the 'Poly' tool, so that the line of the connective WM arch is continued down to the inferior GM of the folia (yellow segmentation in Figure 23-8).

17. When central GM appears be sure to include this and only remove the WM of the brainstem (Figure 23-9).

- The top of the brainstem is clearer on some scans than others. Segment as closely as possible to the GM/WM boundary.
- At this point it is no longer necessary to use the WM segmentation rule shown in Figure 23-8.

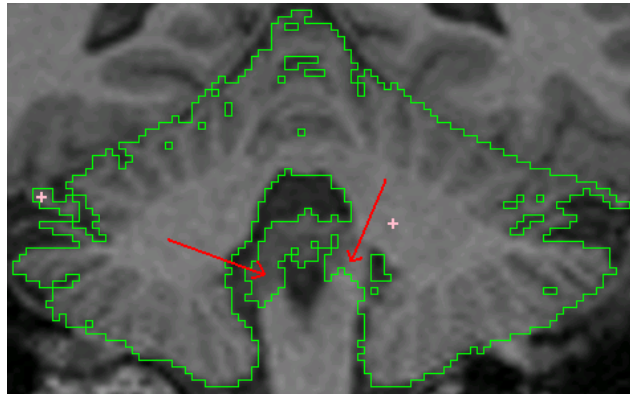


Figure 23-9. Include central GM (red arrows), only removing the WM of the brainstem.

18. The connective tissue between the posterior lobes of the cerebellum should be removed (Figure 23-10). It may also be necessary to add seeds to the posterior folds of the cerebellum.

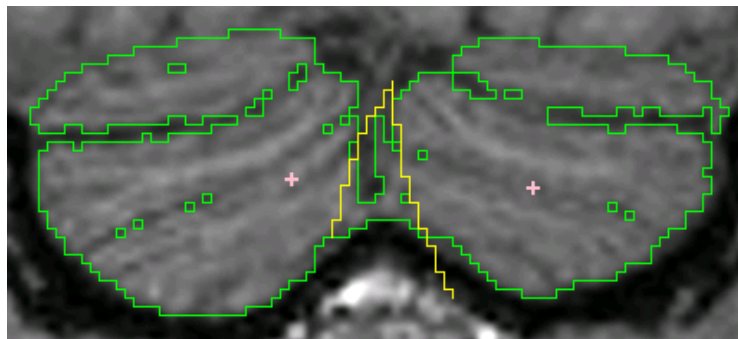


Figure 23-10. Remove the connective tissue between the posterior lobes.

19. Check the segmentation in the sagittal view, making adjustments in the coronal view where necessary.
 - It is best to do as few edits as possible.
20. Save the segmentation; click OK in the editor window, in the main window highlight the region, *Regions>Database In>e.g. Cerebellum*.

23.5.2 Protocol 2

8. In the coronal view seed the cerebellum.
9. Manually remove the brainstem at a low section and separate the cerebellum from the cerebrum, tracing the divide between the two.
 - Use the position bar and the sagittal view if unclear.
 - It may be necessary to seed additional areas at the anterior and posterior ends of the cerebellum.
10. Switch to the sagittal view.
11. Use the poly tool to remove the brain stem by applying a cut-off running from the most anterior GM of the superior cerebellar cortex to the most anterior inferior GM of the cerebellar cortex, e.g. in Figure 23-11.
 - Applying this cut-off in successive slices should produce a smooth divide down the WM in the coronal and axial views.
12. Tidy additional sections in the sagittal view as you go (always referring to the coronal view).

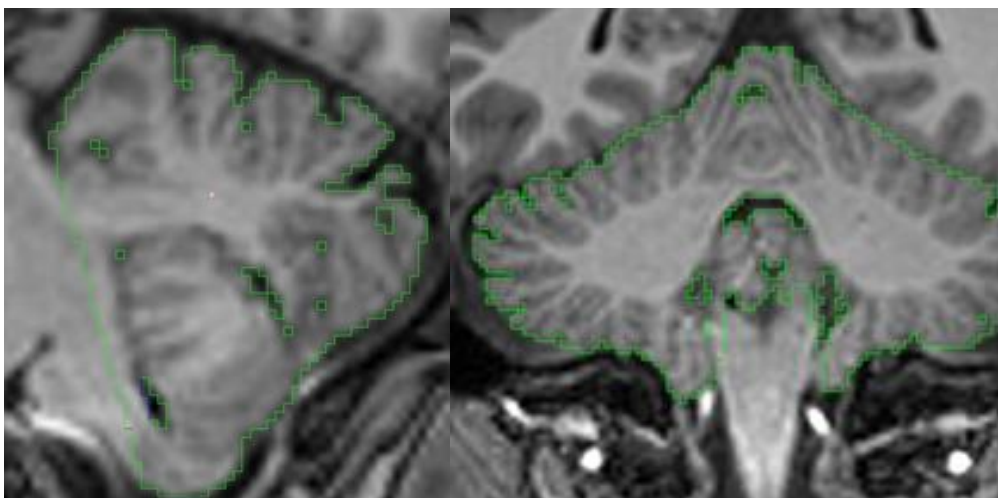


Figure 23-11. Example of a cerebellum segmentation in MIDAS software.

13. Check axial and coronal views applying minor edits.
14. Save in Regions with lower and upper threshold values (e.g. 159_424).

15. Save the segmentation; click OK in the editor window, in the main window highlight the region,
Regions>Database In>e.g. Cerebellum.

References

1996. Unified Huntington's Disease Rating Scale: reliability and consistency. Huntington Study Group. *Movement Disorders*, 11, (2) 136-142
- Adolphs, R. 2002. Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, 12, (2) 169-177
- Alfano, B., Brunetti, A., Covelli, E.M., Quarantelli, M., Panico, M.R., Ciarmiello, A., & Salvatore, M. 1997. Unsupervised, automated segmentation of the normal brain using a multispectral relaxometric magnetic resonance approach. *Magnetic Resonance in Medicine*, 37, (1) 84-93
- Allen, G. & Courchesne, E. 1998. The cerebellum and non-motor function: clinical implications. *Molecular Psychiatry*, 3, (3) 207-210
- Alvarez, J.A. & Emory, E. 2006. Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review*, 16, (1) 17-42
- Andre, R., Scahill, R.I., Haider, S., & Tabrizi, S.J. 2014. Biomarker development for Huntington's disease. *Drug Discovery Today*
- Andrew, S.E., Goldberg, Y.P., Kremer, B., Telenius, H., Theilmann, J., Adam, S., Starr, E., Squitieri, F., Lin, B., Kalchman, M.A., Graham, R.K., & Hayden, M.R. 1993. The Relationship Between Trinucleotide (Cag) Repeat Length and Clinical-Features of Huntington Disease. *American Journal of Human Genetics*, 53, (3) 1118
- Ashburner, J. 2007. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38, (1) 95-113
- Ashburner, J. & Friston, K.J. 2000. Voxel-based morphometry - The methods. *Neuroimage*, 11, (6) 805-821
- Ashburner, J. & Friston, K.J. 2005. Unified segmentation. *Neuroimage*, 26, (3) 839-851
- Atkinson, A.J., Colburn, W.A., DeGruttola, V.G., Demets, D.L., Downing, G.J., Hoth, D.F., Oates, J.A., Peck, C.C., Schooley, R.T., Spilker, B.A., Woodcock, J., & Zeger, S.L. 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, 69, (3) 89-95
- Aylward, E.H., Codori, A.M., Barta, P.E., Pearlson, G.D., Harris, G.J., & Brandt, J. 1996. Basal ganglia volume and proximity to onset in presymptomatic Huntington disease. *Archives of Neurology*, 53, (12) 1293-1296
- Aylward, E.H., Codori, A.M., Rosenblatt, A., Sherr, M., Brandt, J., Stine, O.C., Barta, P.E., Pearlson, G.D., & Ross, C.A. 2000. Rate of caudate atrophy in presymptomatic and symptomatic stages of Huntington's disease. *Movement Disorders*, 15, (3) 552-560
- Aylward, E.H., Li, Q., Stine, O.C., Ranen, N., Sherr, M., Barta, P.E., Bylsma, F.W., Pearlson, G.D., & Ross, C.A. 1997. Longitudinal change in basal ganglia volume in patients with Huntington's disease. *Neurology*, 48, (2) 394-399
- Aylward, E.H., Nopoulos, P.C., Ross, C.A., Langbehn, D.R., Pierson, R.K., Mills, J.A., Johnson, H.J., Magnotta, V.A., Juhl, A.R., & Paulsen, J.S. 2011. Longitudinal change in regional brain volumes in prodromal Huntington disease. *Journal of Neurology, Neurosurgery and Psychiatry*, 82, (4) 405-410
- Aylward, E.H., Rosenblatt, A., Field, K., Yallapragada, V., Kiebertz, K., McDermott, M., Raymond, L.A., Almquist, E.W., Hayden, M., & Ross, C.A. 2003. Caudate volume as an outcome measure in clinical trials for Huntington's disease: a pilot study. *Brain Research Bulletin*, 62, (2) 137-141

- Aylward, E.H., Sparks, B.F., Field, K.M., Yallapragada, V., Shpritz, B.D., Rosenblatt, A., Brandt, J., Gourley, L.M., Liang, K., Zhou, H., Margolis, R.L., & Ross, C.A. 2004. Onset and rate of striatal atrophy in preclinical Huntington disease. *Neurology*, 63, (1) 66-72
- Balcer, L.J., Baier, M.L., Pelak, V.S., Fox, R.J., Shuwairi, S., Galetta, S.L., Cutter, G.R., & Maguire, M.G. 2000. New low-contrast vision charts: reliability and test characteristics in patients with multiple sclerosis. *Multiple Sclerosis*, 6, (3) 163-171
- Barnes, J., Ridgway, G.R., Bartlett, J., Henley, S.M., Lehmann, M., Hobbs, N., Clarkson, M.J., MacManus, D.G., Ourselin, S., & Fox, N.C. 2010. Head size, age and gender adjustment in MRI studies: a necessary nuisance? *Neuroimage*, 53, (4) 1244-1255
- Beauchamp, M.S. 2005. See me, hear me touch me: Multisensory intergration in lateral occipital-temporal cortex. *Current Opinion in Neurobiology*, 15, (2) 145-153
- Bechtel, N., Scahill, R.I., Rosas, H.D., Acharya, T., van den Bogaard, S.J., Jauffret, C., Say, M.J., Sturrock, A., Johnson, H., Onorato, C.E., Salat, D.H., Durr, A., Leavitt, B.R., Roos, R.A., Landwehrmeyer, G.B., Langbehn, D.R., Stout, J.C., Tabrizi, S.J., & Reilmann, R. 2010. Tapping linked to function and structure in premanifest and symptomatic Huntington disease. *Neurology*, 75, (24) 2150-2160
- Beglinger, L.J., Adams, W.H., Langbehn, D., Fiedorowicz, J.G., Jorge, R., Biglan, K., Caviness, J., Olson, B., Robinson, R.G., Kieburtz, K., & Paulsen, J.S. 2014. Results of the Citalopram to Enhance Cognition in Huntington Disease Trial. *Movement Disorders*, 29, 401-405
- Benton, A.L. 1980. The Neuropsychology of Facial Recognition. *American Psychologist*, 35, (2) 176-186
- Bernard, J.A. & Seidler, R.D. 2013. Relationships Between Regional Cerebellar Volume and Sensorimotor and Cognitive Function in Young and Older Adults. *Cerebellum*, 12, (5) 721-737
- Bostan, A.C. & Strick, P.L. 2010. The Cerebellum and Basal Ganglia are Interconnected. *Neuropsychology Review*, 20, (3) 261-270
- Boyes, R.G., Gunter, J.L., Frost, C., Janke, A.L., Yeatman, T., Hill, D.L., Bernstein, M.A., Thompson, P.M., Weiner, M.W., Schuff, N., Alexander, G.E., Killiany, R.J., DeCarli, C., Jack, C.R., & Fox, N.C. 2008. Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. *Neuroimage*, 39, (4) 1752-1762
- Brinkman, R.R., Mezei, M.M., Theilmann, J., Almqvist, E., & Hayden, M.R. 1997. The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. *American Journal of Human Genetics*, 60, (5) 1202-1210
- Buxton, R.B. 2002. *Introduction to functional magnetic resonance imaging: principles and techniques* Cambridge, Cambridge University Press.
- Cabeza, R. & Nyberg, L. 2000. Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, 12, (1) 1-47
- Calder, A.J., Keane, J., Young, A.W., Lawrence, A.D., Mason, S., & Barker, R.A. 2010. The relation between anger and different forms of disgust: implications for emotion recognition impairments in Huntington's disease. *Neuropsychologia*, 48, (9) 2719-2729
- Calder, A.J., Young, A.W., Rowland, D., Perrett, D.I., Hodges, J.R., & Etcoff, N.L. 1996. Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology*, 13, (5) 699-745

- Cardoso, M.J., Clarkson, M.J., Ridgway, G.R., Modat, M., Fox, N.C., & Ourselin, S. 2011. LoAd: A locally adaptive cortical segmentation algorithm. *Neuroimage*, 56, (3) 1386-1397
- Cardoso, M.J., Leung, K., Modat, M., Keihaninejad, S., Cash, D., Barnes, J., Fox, N.C., & Ourselin, S. 2013. STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcellation. *Medical Image Analysis*, 17, (6) 671-684
- Carpenter, J. & Bithell, J. 2000. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, 19, (9) 1141-1164
- Christensen, G.E., Rabbitt, R.D., & Miller, M.I. 1996. Deformable templates using large deformation kinematics. *Ieee Transactions on Image Processing*, 5, (10) 1435-1447
- Ciarmiello, A., Cannella, M., Lastoria, S., Simonelli, M., Frati, L., Rubinsztein, D.C., & Squitieri, F. 2006. Brain white-matter volume loss and glucose hypometabolism precede the clinical symptoms of Huntington's disease. *Journal of Nuclear Medicine*, 47, (2) 215-222
- Cole, J., Chaddock, C.A., Farmer, A.E., Aitchison, K.J., Simmons, A., McGuffin, P., & Fu, C.H.Y. 2012. White matter abnormalities and illness severity in major depressive disorder. *British Journal of Psychiatry*, 201, (1) 33-39
- Cowan, N., Elliott, E.M., Scott, S.J., Morey, C.C., Mattox, S., Hismjatullina, A., & Conway, A.R. 2005. On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, 51, (1) 42-100
- Craufurd, D. & Snowden, J. 2002, "Neuropsychological and neuropsychiatric aspects of Huntington's Disease," Oxford: Oxford University Press, pp. 62-94.
- Crum, W.R., Hartkens, T., & Hill, D.L.G. 2004. Non-rigid image registration: theory and practice. *British Journal of Radiology*, 77, S140-S153
- D'Angelo, E. 2010. Neuronal circuit function and dysfunction in the cerebellum: from neurons to integrated control. *Functional Neurology*, 25, (3) 125-127
- D'Angelo, E. 2011. Neural Circuits of the Cerebellum: Hypothesis for Function. *Journal of Integrative Neuroscience*, 10, (3) 317-352
- Dale, A.M., Fischl, B., & Sereno, M.I. 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*, 9, (2) 179-194
- de Boo, G., Tibben, A., Hermans, J., Maat, A., & Roos, R.A.C. 1998. Subtle involuntary movements are not reliable indicators of incipient Huntington's disease. *Movement Disorders*, 13, (1) 96-99
- de la Monte, S.M., Vonsattel, J.P., & Richardson, E.P., Jr. 1988. Morphometric demonstration of atrophic changes in the cerebral cortex, white matter, and neostriatum in Huntington's disease. *Journal of Neuropathology and Experimental Neurology*, 47, (5) 516-525
- deGelder, B., Van den Stock, J., Balaguer, R.D., & Bachoud-Levi, A.C. 2008. Huntington's disease impairs recognition of angry and instrumental body language. *Neuropsychologia*, 46, (1) 369-373
- Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., & Killiany, R.J. 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, 31, (3) 968-980

- Devinsky, O., Morrell, M.J., & Vogt, B.A. 1995. Contributions of Anterior Cingulate Cortex to Behaviour. *Brain*, 118, 279-306
- Djousse, L., Knowlton, B., Hayden, M., Almquist, E.W., Brinkman, R., Ross, C., Margolis, R., Rosenblatt, A., Durr, A., Dode, C., Morrison, P.J., Novelletto, A., Frontali, M., Trent, R.J.A., McCusker, E., Gomez-Tortosa, E., Mayo, D., Jones, R., Zanko, A., Nance, M., Abramson, R., Suchowersky, O., Paulsen, J., Harrison, M., Yang, Q., Cupples, L.A., Gusella, J.F., MacDonald, M.E., & Myers, R.H. 2003. Interaction of normal and expanded CAG repeat sizes influences age at onset of Huntington disease. *American Journal of Medical Genetics Part A*, 119A, (3) 279-282
- Dominguez, J.F., Egan, G.F., Gray, M.A., Poudel, G.R., Churchyard, A., Chua, P., Stout, J.C., & Georgiou-Karistianis, N. 2013. Multi-Modal Neuroimaging in Premanifest and Early Huntington's Disease: 18 Month Longitudinal Data from the IMAGE-HD Study. *Plos One*, 8, (9)
- Dumas, E.M., van den Bogaard, S.J.A., Ruber, M.E., Reilman, R.R., Stout, J.C., Craufurd, D., Hicks, S.L., Kennard, C., Tabrizi, S.J., van Buchem, M.A., van der Grond, J., & Roos, R.A.C. 2012. Early changes in white matter pathways of the sensorimotor cortex in premanifest Huntington's disease. *Human Brain Mapping*, 33, (1) 203-212
- Durr, A. 2010. Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond. *Lancet Neurology*, 9, (9) 885-894
- Efron, B. & Tibshirani, R.J. 1993. *An introduction to the bootstrap* New York, Chapman & Hall.
- Ekman, P. 1992. Are There Basic Emotions. *Psychological Review*, 99, (3) 550-553
- Evans, A., Collins, D., Mills, S., Brown, E., Kelly, R., & Peters, T. 3D statistical neuroanatomical models from 305 MRI volumes, pp. 1813-1817.
- Feigin, A., Tang, C., Ma, Y., Mattis, P., Zgaljardic, D., Guttman, M., Paulsen, J.S., Dhawan, V., & Eidelberg, D. 2007. Thalamic metabolism and symptom onset in preclinical Huntington's disease. *Brain*, 130, (Pt 11) 2858-2867
- Fennema-Notestine, C., Archibald, S.L., Jacobson, M.W., Corey-Bloom, J., Paulsen, J.S., Peavy, G.M., Gamst, A.C., Hamilton, J.M., Salmon, D.P., & Jernigan, T.L. 2004. In vivo evidence of cerebellar atrophy and cerebral white matter loss in Huntington disease. *Neurology*, 63, (6) 989-995
- Fischl, B. & Dale, A.M. 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences U.S.A*, 97, (20) 11050-11055
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A.M. 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33, (3) 341-355
- Fischl, B., Sereno, M.I., & Dale, A.M. 1999. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9, (2) 195-207
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., & Dale, A.M. 2004. Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14, (1) 11-22
- Fitzpatrick, L.E. & Crowe, S.F. 2013. Cognitive and emotional deficits in chronic alcoholics: a role for the cerebellum? *Cerebellum*, 12, (4) 520-533

- Folstein, S.E., Jensen, B., Leigh, R.J., & Folstein, M.F. 1983. The measurement of abnormal movement: methods developed for Huntington's disease. *Neurobehavioral Toxicology and Teratology*, 5, (6) 605-609
- Freeborough, P.A. & Fox, N.C. 1997. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Transactions on Medical Imaging*, 16, (5) 623-629
- Freeborough, P.A. & Fox, N.C. 1998. Modeling brain deformations in Alzheimer disease by fluid registration of serial 3D MR images. *Journal of Computer Assisted Tomography*, 22, (5) 838-843
- Freeborough, P.A., Fox, N.C., & Kitney, R.I. 1997. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. *Computer Methods and Programs in Biomedicine*, 53, (1) 15-25
- Gegenfurtner, K.R., Kiper, D.C., & Levitt, J.B. 1997. Functional properties of neurons in macaque area V3. *Journal of Neurophysiology*, 77, (4) 1906-1923
- Genovese, C.R., Lazar, N.A., & Nichols, T. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15, (4) 870-878
- Georgiou-Karistianis, N., Poudel, G.R., Dominguez, D.J.F., Langmaid, R., Gray, M.A., Churchyard, A., Chua, P., Borowsky, B., Egan, G.F., & Stout, J.C. 2013a. Functional and connectivity changes during working memory in Huntington's disease: 18 month longitudinal data from the IMAGE-HD study. *Brain and Cognition*, 83, (1) 80-91
- Georgiou-Karistianis, N., Scahill, R., Tabrizi, S.J., Squitieri, F., & Aylward, E. 2013b. Structural MRI in Huntington's disease and recommendations for its potential use in clinical trials. *Neuroscience and Biobehavioral Reviews*, 37, (3) 480-490
- Gilman, S., Koller, M., Black, R.S., Jenkins, L., Griffith, S.G., Fox, N.C., Eisner, L., Kirby, L., Rovira, M.B., Forette, F., & Orgogozo, J.M. 2005. Clinical effects of A beta immunization (AN1792) in patients with AD in an interrupted trial. *Neurology*, 64, (9) 1553-1562
- Gomez-Anson, B., Alegret, M., Munoz, E., Monte, G.C., Alayrach, E., Sanchez, A., Boada, M., & Tolosa, E. 2009. Prefrontal cortex volume reduction on MRI in preclinical Huntington's disease relates to visuomotor performance and CAG number. *Parkinsonism & Related Disorders*, 15, (3) 213-219
- Gray, J.M., Young, A.W., Barker, W.A., Curtis, A., & Gibson, D. 1997. Impaired recognition of disgust in Huntington's disease gene carriers. *Brain*, 120 (Pt 11), 2029-2038
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. 2001. The lateral occipital complex and its role in object recognition. *Vision Research*, 41, (10-11) 1409-1422
- Gronenschild, E.H., Habets, P., Jacobs, H.I., Mengelers, R., Rozendaal, N., van, O.J., & Marcelis, M. 2012. The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements. *Plos One*, 7, (6) e38234
- Hall, J. A. 2001, "The PONS test and the psychometric approach to measuring interpersonal sensitivity," *In Interpersonal Sensitivity: Theory and Measurement*, J. A. Hall & F. J. Bernieri, eds., Mahwah, NJ: Erlbaum, pp. 143-160.
- Halliday, G.M., McRitchie, D.A., Macdonald, V., Double, K.L., Trent, R.J., & McCusker, E. 1998. Regional specificity of brain atrophy in Huntington's disease. *Experimental Neurology*, 154, (2) 663-672

- Handley, O.J., Naji, J.J., Dunnett, S.B., & Rosser, A.E. 2006. Pharmaceutical, cellular and genetic therapies for Huntington's disease. *Clinical Science*, 110, (1) 73-88
- Hart, E., Middelkoop, H., Jurgens, C.K., Witjes-Ane, M.N., & Roos, R.A. 2011. Seven-year clinical follow-up of premanifest carriers of Huntington's disease. *PLoS Currents*, 3, RRN1288
- Hayes, C.J., Stevenson, R.J., & Coltheart, M. 2007. Disgust and Huntington's disease. *Neuropsychologia*, 45, (6) 1135-1151
- Hayes, C.J., Stevenson, R.J., & Coltheart, M. 2009. The Processing of Emotion in Patients With Huntington's Disease: Variability and Differential Deficits in Disgust. *Cognitive and Behavioral Neurology*, 22, (4) 249-257
- Henley, S.M., Frost, C., MacManus, D.G., Warner, T.T., Fox, N.C., & Tabrizi, S.J. 2006. Increased rate of whole-brain atrophy over 6 months in early Huntington disease. *Neurology*, 67, (4) 694-696
- Henley, S.M., Novak, M.J., Frost, C., King, J., Tabrizi, S.J., & Warren, J.D. 2012. Emotion recognition in Huntington's disease: a systematic review. *Neuroscience and Biobehavioral Reviews*, 36, (1) 237-253
- Henley, S.M., Ridgway, G.R., Scahill, R.I., Kloppel, S., Tabrizi, S.J., Fox, N.C., & Kassubek, J. 2010. Pitfalls in the use of voxel-based morphometry as a biomarker: examples from huntington disease. *AJNR American Journal of Neuroradiology*, 31, (4) 711-719
- Henley, S.M., Wild, E.J., Hobbs, N.Z., Frost, C., MacManus, D.G., Barker, R.A., Fox, N.C., & Tabrizi, S.J. 2009. Whole-brain atrophy as a measure of progression in premanifest and early Huntington's disease. *Movement Disorders*, 24, (6) 932-936
- Henley, S.M., Wild, E.J., Hobbs, N.Z., Warren, J.D., Frost, C., Scahill, R.I., Ridgway, G.R., MacManus, D.G., Barker, R.A., Fox, N.C., & Tabrizi, S.J. 2008. Defective emotion recognition in early HD is neuropsychologically and anatomically generic. *Neuropsychologia*, 46, (8) 2152-2160
- Hjermind, L.E., Law, I., Jonch, A., Stokholm, J., & Nielsen, J.E. 2011. Huntington's disease: effect of memantine on FDG-PET brain metabolism? *Journal of Neuropsychiatry and Clinical Neurosciences*, 23, (2) 206-210
- Hobbs, N.Z., Barnes, J., Frost, C., Henley, S.M., Wild, E.J., Macdonald, K., Barker, R.A., Scahill, R.I., Fox, N.C., & Tabrizi, S.J. 2010a. Onset and progression of pathologic atrophy in Huntington disease: a longitudinal MR imaging study. *AJNR American Journal of Neuroradiology*, 31, (6) 1036-1041
- Hobbs, N.Z., Cole, J.H., Farmer, R.E., Rees, E.M., Crawford, H.E., Malone, I.B., Roos, R.A., Sprengelmeyer, R., Durr, A., Landwehrmeyer, B., Scahill, R.I., Tabrizi, S.J., & Frost, C. 2013. Evaluation of multi-modal, multi-site neuroimaging measures in Huntington's disease: Baseline results from the PADDINGTON study. *Neuroimage: Clinical*, 2, 204-211
- Hobbs, N.Z., Henley, S.M., Ridgway, G.R., Wild, E.J., Barker, R.A., Scahill, R.I., Barnes, J., Fox, N.C., & Tabrizi, S.J. 2010b. The progression of regional atrophy in premanifest and early Huntington's disease: a longitudinal voxel-based morphometry study. *Journal of Neurology, Neurosurgery and Psychiatry*, 81, (7) 756-763
- Hobbs, N.Z., Henley, S.M.D., Wild, E.J., Leung, K.K., Frost, C., Barker, R.A., Scahill, R.I., Barnes, J., Tabrizi, S.J., & Fox, N.C. 2009. Automated quantification of caudate atrophy by local registration of serial MRI: Evaluation and application in Huntington's disease. *Neuroimage*, 47, (4) 1659-1665

- Holland, D., Brewer, J.B., Hagler, D.J., Fennema-Notestine, C., & Dale, A.M. 2009. Subregional neuroanatomical change as a biomarker for Alzheimer's disease. *Proceedings of the National Academy of Science U.S.A*, 106, (49) 20954-20959
- Holland, D. & Dale, A.M. 2011. Nonlinear registration of longitudinal images and measurement of change in regions of interest. *Medical Image Analysis*, 15, (4) 489-497
- Holroyd, S., Shepherd, M.L., & Downs, J.H., III 2000. Occipital atrophy is associated with visual hallucinations in Alzheimer's disease. *Journal of Neuropsychiatry and Clinical Neuroscience*, 12, (1) 25-28
- HORIZON Investigators of the Huntington Study Group and European Huntington's Disease Network. 2013. A randomized, double-blind, placebo-controlled study of latrepirdine in patients with mild to moderate Huntington disease. *JAMA Neurology*, 70, (1) 25-33
- Huntington Study Group DOMINO Investigators 2010. A futility study of minocycline in Huntington's disease. *Movement Disorders*, 25, (13) 2219-2224
- Huntington Study Group TREND-HD Investigators 2008. Randomized controlled trial of ethyl-eicosapentaenoic acid in Huntington disease: the TREND-HD study. *Archives of Neurology*, 65, (12) 1582-1589
- Huntington's Disease Collaborative Research Group 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's Disease chromosomes. *Cell*, 72, 971-983
- Hyson, H.C., Kieburtz, K., Shoulson, I., McDermott, M., Ravina, B., de Bleeck, E.A., Cudkowicz, M.E., Ferrante, R.J., Como, P., Frank, S., Zimmerman, C., Cudkowicz, M.E., Ferrante, K., Newhall, K., Jennings, D., Kelsey, T., Walker, F., Hunt, V., Daigneault, S., Goldstein, M., Weber, J., Watts, A., Beal, M.F., Browne, S.E., & Metakis, L.J. 2010. Safety and Tolerability of High-Dosage Coenzyme Q(10) in Huntington's Disease and Healthy Subjects. *Movement Disorders*, 25, (12) 1924-1928
- Ille, R., Holl, A.K., Kapfhammer, H.P., Reisinger, K., Schafer, A., & Schienle, A. 2011a. Emotion recognition and experience in Huntington's disease: is there a differential impairment? *Psychiatry Research*, 188, (3) 377-382
- Ille, R., Schafer, A., Scharmuller, W., Enzinger, C., Schoggl, H., Kapfhammer, H.P., & Schienle, A. 2011b. Emotion recognition and experience in Huntington disease: a voxel-based morphometry study. *Journal of Psychiatry and Neuroscience*, 36, (6) 383-390
- Jack, C.R., Petersen, R.C., Xu, Y.C., Waring, S.C., O'Brien, P.C., Tangalos, E.G., Smith, G.E., Ivnik, R.J., & Kokmen, E. 1997. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology*, 49, (3) 786-794
- Jankovic, J., Vuong, K.D., & Hunter, C. 2004. Safety and efficacy of tetrabenazine in the treatment of Huntington's disease and other choreas. *Movement Disorders*, 19, (9) 1127
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17, (2) 825-841
- Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., & Smith, S.M. 2012. Fsl. *Neuroimage*, 62, (2) 782-790
- Jeste, D.V., Barban, L., & Parisi, J. 1984. Reduced Purkinje cell density in Huntington's disease. *Experimental Neurology*, 85, (1) 78-86

- Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., Maguire, P., Rosas, D., Makris, N., Gollub, R., Dale, A., Dickerson, B.C., & Fischl, B. 2009. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *Neuroimage*, 46, (1) 177-192
- Julious, S.A. 2009. *Sample sizes in clinical trials* Boca Raton: Chapman and Hall.
- Kanwisher, N., McDermott, J., & Chun, M.M. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17, (11) 4302-4311
- Kiebertz, K., MacDonald, M., Shih, C., Feigin, A., Steinberg, K., Bordwell, K., Zimmerman, C., Srinidhi, J., Sotack, J., Gusella, J., & . 1994. Trinucleotide repeat length and progression of illness in Huntington's disease. *Journal of Medical Genetics*, 31, (11) 872-874
- Kiebertz, K., McDermott, M.P., Voss, T.S., Corey-Bloom, J., Deuel, L.M., Dorsey, E.R., Factor, S., Geschwind, M.D., Hodgeman, K., Kayson, E., Noonberg, S., Pourfar, M., Rabinowitz, K., Ravina, B., Sanchez-Ramos, J., Seely, L., Walker, F., & Feigin, A. 2010. A randomized, placebo-controlled trial of latrepirdine in Huntington disease. *Archives of Neurology*, 67, (2) 154-160
- Kim, J.W., Lee, D.Y., Choo, I.H., Seo, E.H., Kim, S.G., Park, S.Y., & Woo, J.I. 2011. Microstructural Alteration of the Anterior Cingulum is Associated With Apathy in Alzheimer Disease. *American Journal of Geriatric Psychiatry*, 19, (7) 644-653
- Kipps, C.M., Duggins, A.J., Mahant, N., Gomes, L., Ashburner, J., & McCusker, E.A. 2005. Progression of structural neuropathology in preclinical Huntington's disease: a tensor based morphometry study. *Journal of Neurology, Neurosurgery and Psychiatry*, 76, (5) 650-655
- Kipps, C.M., Duggins, A.J., McCusker, E.A., & Calder, A.J. 2007. Disgust and happiness recognition correlate with anteroventral insula and amygdala volume respectively in preclinical Huntington's disease. *Journal of Cognitive Neuroscience*, 19, (7) 1206-1217
- Kloppel, S., Draganski, B., Golding, C.V., Chu, C., Nagy, Z., Cook, P.A., Hicks, S.L., Kennard, C., Alexander, D.C., Parker, G.J., Tabrizi, S.J., & Frackowiak, R.S. 2008. White matter connections reflect changes in voluntary-guided saccades in pre-symptomatic Huntington's disease. *Brain*, 131, (Pt 1) 196-204
- Kremer, B., Clark, C.M., Almqvist, E.W., Raymond, L.A., Graf, P., Jacova, C., Mezei, M., Hardy, M.A., Snow, B., Martin, W., & Hayden, M.R. 1999. Influence of lamotrigine on progression of early Huntington disease: a randomized clinical trial. *Neurology*, 53, (5) 1000-1011
- Kunimatsu, N., Aoki, S., Kunimatsu, A., Yoshida, M., Abe, O., Yamada, H., Masutani, Y., Kasai, K., Yamasue, H., Ohtsu, H., & Ohtomo, K. 2008. Tract-specific analysis of the superior occipitofrontal fasciculus in schizophrenia. *Psychiatry Research-Neuroimaging*, 164, (3) 198-205
- Kutty, I.N. & Prendes, J.L. 1981. Psychosis and Cerebellar Degeneration. *Journal of Nervous and Mental Disease*, 169, (6) 390-391
- Labuschagne, I., Jones, R., Callaghan, J., Whitehead, D., Dumas, E.M., Say, M.J., Hart, E.P., Justo, D., Coleman, A., Santos, R.C.D., Frost, C., Craufurd, D., Tabrizi, S.J., & Stout, J.C. 2013. Emotional face recognition deficits and medication effects in pre-manifest through stage-II Huntington's disease. *Psychiatry Research*, 207, (1-2) 118-126
- Lancet Editorial 2010. Dispelling the stigma of Huntington's disease. *Lancet Neurology*, 9, (8) 751

- Landwehrmeyer, G.B., Dubois, B., de Yebenes, J.G., Kremer, B., Gaus, W., Kraus, P.H., Przuntek, H., Dib, M., Doble, A., Fischer, W., & Ludolph, A.C. 2007. Riluzole in Huntington's disease: A 3-year, randomized controlled study. *Annals of Neurology*, 62, (3) 262-272
- Langbehn, D.R., Brinkman, R.R., Falush, D., Paulsen, J.S., & Hayden, M.R. 2004. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. *Clinical Genetics*, 65, (4) 267-277
- Langbehn, D.R., Hayden, M.R., & Paulsen, J.S. 2010. CAG-repeat length and the age of onset in Huntington disease (HD): a review and validation study of statistical approaches. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 153B, (2) 397-408
- Lange, H., Thorner, G., Hopf, A., & Schroder, K.F. 1976. Morphometric studies of the neuropathological changes in choreatic diseases. *Journal of Neurological Science*, 28, (4) 401-425
- Larsson, J. & Heeger, D.J. 2006. Two retinotopic visual areas in human lateral occipital cortex. *Journal of Neuroscience*, 26, (51) 13128-13142
- Lauterbach, E.C. 1996. Bipolar disorders, dystonia, and compulsion after dysfunction of the cerebellum, dentatorubrothalamic tract, and substantia nigra. *Biological Psychiatry*, 40, (8) 726-730
- Leung, K.K., Clarkson, M.J., Bartlett, J.W., Clegg, S., Jack, C.R., Jr., Weiner, M.W., Fox, N.C., & Ourselin, S. 2010. Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection. *Neuroimage*, 50, (2) 516-523
- Leung, K.K., Ridgway, G.R., Ourselin, S., & Fox, N.C. 2012. Consistent multi-time-point brain atrophy estimation from the boundary shift integral. *Neuroimage*, 59, (4) 3995-4005
- Lewis, E.B. & Fox, N.C. 2004. Correction of differential intensity inhomogeneity in longitudinal MR images. *Neuroimage*, 23, (1) 75-83
- Li, J.L., Hayden, M.R., Almqvist, E.W., Brinkman, R.R., Durr, A., Dode, C., Morrison, P.J., Suchowersky, O., Ross, C.A., Margolis, R.L., Rosenblatt, A., Gomez-Tortosa, E., Cabrero, D.M., Novelletto, A., Frontali, M., Nance, M., Trent, R.J., McCusker, E., Jones, R., Paulsen, J.S., Harrison, M., Zanko, A., Abramson, R.K., Russ, A.L., Knowlton, B., Djousse, L., Mysore, J.S., Tariot, S., Gusella, M.F., Wheeler, V.C., Atwood, L.D., Cupples, L.A., Saint-Hilaire, M., Cha, J.H., Hersch, S.M., Koroshetz, W.J., Gusella, J.F., MacDonald, M.E., & Myers, R.H. 2003. A genome scan for modifiers of age at onset in Huntington disease: The HD MAPS study. *American Journal of Human Genetics*, 73, (3) 682-687
- Lundin, A., Dietrichs, E., Haghighi, S., Goller, M.L., Heiberg, A., Loutfi, G., Widner, H., Wiktorin, K., Wiklund, L., Svenningsson, A., Sonesson, C., Waters, N., Waters, S., & Tedroff, J. 2010. Efficacy and safety of the dopaminergic stabilizer Pridopidine (ACR16) in patients with Huntington's disease. *Clinical Neuropharmacology*, 33, (5) 260-264
- Lyoo, I.K., Sung, Y.H., Dager, S.R., Friedman, S.D., Lee, J.Y., Kim, S.J., Kim, N., Dunner, D.L., & Renshaw, P.F. 2006. Regional cerebral cortical thinning in bipolar disorder. *Bipolar Disorder*, 8, (1) 65-74
- MacLeod, C.M. 1991. Half a century of research on the Stroop effect: an integrative review. *Psychology Bulletin*, 109, (2) 163-203
- Magnotta, V.A., Harris, G., Andreasen, N.C., O'Leary, D.S., Yuh, W.T., & Heckel, D. 2002. Structural MR image processing using the BRAINS2 toolbox. *Computerized Medical Imaging and Graphics*, 26, (4) 251-264

- Maguire, E.A., Gadian, D.G., Johnsrude, I.S., Good, C.D., Ashburner, J., Frackowiak, R.S.J., & Frith, C.D. 2000. Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America*, 97, (8) 4398-4403
- Majid, D.S., Aron, A.R., Thompson, W., Sheldon, S., Hamza, S., Stoffers, D., Holland, D., Goldstein, J., Corey-Bloom, J., & Dale, A.M. 2011a. Basal ganglia atrophy in prodromal Huntington's disease is detectable over one year using automated segmentation. *Movement Disorders*.
- Majid, D.S., Stoffers, D., Sheldon, S., Hamza, S., Thompson, W.K., Goldstein, J., Corey-Bloom, J., & Aron, A.R. 2011b. Automated structural imaging analysis detects premanifest Huntington's disease neurodegeneration within 1 year. *Movement Disorders*, 26, (8) 1481-1488
- Manto, M. 2012. Toxic agents causing cerebellar ataxias. *Handbook of Clinical Neurology*, 103, 201-213
- Maucksch, C., Vazey, E.M., Gordon, R.J., & Connor, B. 2013. Stem cell-based therapy for Huntington's disease. *Journal of Cellular Biochemistry*, 114, (4) 754-763
- Mechelli, A., Humphreys, G.W., Mayall, K., Olson, A., & Price, C.J. 2000. Differential effects of word length and visual contrast in the fusiform and lingual gyri during reading. *Proceedings of the Royal Society B: Biological Sciences*, 267, (1455) 1909-1913
- Mejia, N.I., Hunter, C.B., Flores, K., & Ondo, W.G. 2005. A pilot study of the clinical efficacy and safety of memantine for Huntington's disease. *Movement Disorders*, 20, S10-S11
- Mickes, L., Jacobson, M., Peavy, G., Wixted, J.T., Lessig, S., Goldstein, J.L., & Corey-Bloom, J. 2010. A Comparison of Two Brief Screening Measures of Cognitive Impairment in Huntington's Disease. *Movement Disorders*, 25, (13) 2229-2233
- Milders, M., Crawford, J.R., Lamb, A., & Simpson, S.A. 2003. Differential deficits in expression recognition in gene-carriers and patients with Huntington's disease. *Neuropsychologia*, 41, (11) 1484-1492
- Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., & Ourselin, S. 2010. Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine*, 98, (3) 278-284
- Montagne, B., Kessels, R.P., Kammers, M.P., Kingma, E., De Haan, E.H., Roos, R.A., & Middelkoop, H.A. 2006. Perception of emotional facial expressions at different intensities in early-symptomatic Huntington's disease. *European Neurology*, 55, (3) 151-154
- Morrison, P.J. 2010. Accurate prevalence and uptake of testing for Huntington's disease. *Lancet Neurology*, 9, (12) 1147
- Morrison, P.J., Harding-Lester, S., & Bradley, A. 2011. Uptake of Huntington disease predictive testing in a complete population. *Clinical Genetics*, 80, (3) 281-286
- Muller, H.P., Gron, G., Sprengelmeyer, R., Kassubek, J., Ludolph, A.C., Hobbs, N., Cole, J., Roos, R.A., Duerr, A., Tabrizi, S.J., Landwehrmeyer, G.B., & Sussmuth, S.D. 2013. Evaluating multicenter DTI data in Huntington's disease on site specific effects: An ex post facto approach. *Neuroimage Clinical*, 2, 161-167
- Muller, R.H., Courchesne, E., & Allen, G. 1998. The cerebellum: So much more. *Science*, 282, (5390) 879-880
- Neychev, V.K., Fan, X., Mitev, V.I., Hess, E.J., & Jinnah, H.A. 2008. The basal ganglia and cerebellum interact in the expression of dystonic movement. *Brain*, 131, (Pt 9) 2499-2509

- Niethammer, M., San Jose, E.R., Bouix, S., Shenton, M., & Westin, C.F. 2006. On diffusion tensor estimation. *Conference Proceedings of the IEEE Engineering in Medicine & Biology Society*, 1, 2622-2625
- Nopoulos, P.C., Aylward, E.H., Ross, C.A., Johnson, H.J., Magnotta, V.A., Juhl, A.R., Pierson, R.K., Mills, J., Langbehn, D.R., & Paulsen, J.S. 2010. Cerebral cortex structure in prodromal Huntington disease. *Neurobiology of Disease*, 40, (3) 544-554
- Novak, M.J.U. & Tabrizi, S.J. 2011. Huntington's Disease: Clinical Presentation and Treatment. *Pathophysiology, Pharmacology, and Biochemistry of Dyskinesia*, 98, 297-323
- Ourselin, S., Roche, A., Subsol, G., Pennec, X., & Ayache, N. 2001. Reconstructing a 3D structure from serial histological sections. *Image and Vision Computing*, 19, (1-2) 25-31
- Paganini, M., Biggeri, A., Romoli, A.M., Mechi, C., Ghelli, E., Berti, V., Pradella, S., Bucciantini, S., Catelan, D., Saccardi, R., Lombardini, L., Mascaldi, M., Massacesi, L., Porfirio, B., Di, L.N., Vannelli, G.B., & Gallina, P. 2013. Fetal striatal grafting slows motor and cognitive decline of Huntington's disease. *Journal of Neurology, Neurosurgery and Psychiatry*
- Papoutsis, M., Labuschagne, I., Tabrizi, S.J., & Stout, J.C. 2014. The cognitive burden in Huntington's disease: Pathology, phenotype, and mechanisms of compensation. *Movement Disorders*, 29, (5) 673-683
- Patenaude, B., Smith, S.M., Kennedy, D.N., & Jenkinson, M. 2011. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage*, 56, (3) 907-922
- Paulsen, J.S. 2009. Functional imaging in Huntington's disease. *Experimental Neurology*, 216, (2) 272-277
- Paulsen, J.S., Hayden, M., Stout, J.C., Langbehn, D.R., Aylward, E., Ross, C.A., Guttman, M., Nance, M., Kiebertz, K., Oakes, D., Shoulson, I., Kayson, E., Johnson, S., & Penziner, E. 2006. Preparing for preventive clinical trials: the Predict-HD study. *Archives of Neurology*, 63, (6) 883-890
- Paulsen, J.S., Langbehn, D.R., Stout, J.C., Aylward, E., Ross, C.A., Nance, M., Guttman, M., Johnson, S., MacDonald, M., Beglinger, L.J., Duff, K., Kayson, E., Biglan, K., Shoulson, I., Oakes, D., & Hayden, M. 2008. Detection of Huntington's disease decades before diagnosis: the Predict-HD study. *Journal of Neurology Neurosurgery and Psychiatry*, 79, (8) 874-880
- Paulsen, J.S., Long, J.D., Johnson, H.J., Aylward, E.H., Ross, C.A., Williams, J.K., Nance, M.A., Erwin, C.J., Westervelt, H.J., Harrington, D.L., Bockholt, H.J., Zhang, Y., McCusker, E.A., Chiu, E.M., & Panegyres, P.K. 2014. Clinical and Biomarker Changes in Premanifest Huntington Disease Show Trial Feasibility: A Decade of the PREDICT-HD Study. *Frontiers in Aging Neuroscience*, 6, 78
- Penney, J.B., Vonsattel, J.P., MacDonald, M.E., Gusella, J.F., & Myers, R.H. 1997. CAG repeat number governs the development rate of pathology in Huntington's disease. *Annals of Neurology*, 41, (5) 689-692
- Prudente, C.N., Hess, E.J., & Jinnah, H.A. 2014. Dystonia as a network disorder: what is the role of the cerebellum? *Neuroscience*, 260, 23-35
- Puce, A., Allison, T., Bentin, S., Gore, J.C., & McCarthy, G. 1998. Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience*, 18, (6) 2188-2199
- Puri, B.K., Bydder, G.M., Counsell, S.J., Corridan, B.J., Richardson, A.J., Hajnal, J.V., Appel, C., McKee, H.M., Vaddadi, K.S., & Horrobin, D.F. 2002. MRI and neuropsychological improvement in Huntington disease following ethyl-EPA treatment. *Neuroreport*, 13, (1) 123-126

- Puri, B.K., Bydder, G.M., Manku, M.S., Clarke, A., Waldman, A.D., & Beckmann, C.F. 2008. Reduction in cerebral atrophy associated with ethyl-eicosapentaenoic acid treatment in patients with Huntington's disease. *Journal of International Medical Research*, 36, (5) 896-905
- Ranen, N.G., Stine, O.C., Abbott, M.H., Sherr, M., Codori, A.M., Franz, M.L., Chao, N.I., Chung, A.S., Pleasant, N., Callahan, C., & . 1995. Anticipation and instability of IT-15 (CAG)_n repeats in parent-offspring pairs with Huntington disease. *American Journal of Human Genetics*, 57, (3) 593-602
- Rees, E.M., Scahill, R., & Hobbs, N. 2013. Longitudinal Neuroimaging Biomarkers in Huntington's Disease. *Journal of Huntington's Disease*, 2, (1) 21-39
- Reilmann, R., Kirsten, F., Bohlen, S., Saemann, P., Merl, T., Auer, D., Klopstock, T., Bender, A., Dose, M., Weindel, A., & Ringelstein, E.B. 2005. Multimodal objective assessment of motor deficits in Huntington's disease using isometric force analysis. *Journal of Neurology Neurosurgery and Psychiatry*, 76,
- Reitan, R.M. & Wolfson, D. 1994. A Selective and Critical-Review of Neuropsychological Deficits and the Frontal Lobes. *Neuropsychology Review*, 4, (3) 161-198
- Reuter, M. & Fischl, B. 2011. Avoiding asymmetry-induced bias in longitudinal image processing. *Neuroimage*, 57, (1) 19-21
- Reuter, M., Schmansky, N.J., Rosas, H.D., & Fischl, B. 2012. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage*, 61, (4) 1402-1418
- Robertson, I.H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. 1994. *The Test of Everyday Attention* Bury St. Edmunds (UK), Thames Valley Test Co.
- Robotham, L., Sauter, D.A., Bachoud-Levi, A.C., & Trinkler, I. 2011. The impairment of emotion recognition in Huntington's disease extends to positive emotions. *Cortex*, 47, (7) 880-884
- Rodda, R.A. 1981. Cerebellar atrophy in Huntington's disease. *Journal of Neurological Science*, 50, (1) 147-157
- Roe, A.W. & Ts'o, D.Y. 1995. Visual topography in primate V2: multiple representation across functional stripes. *Journal of Neuroscience*, 15, (5 Pt 2) 3689-3715
- Rosas, D., Doros, G., Gevorkian, S., Malarick, K., Reuter, M., Coutu, J.P., Triggs, T., Wilkens, P.J., Matson, W., Salat, D.H., & Hersch, S. 2014. PRECREST: A phase II prevention and biomarker trial of creatine in at-risk Huntington's Disease. *Neurology*, 82, 1-8
- Rosas, H.D., Hevelone, N.D., Zaleta, A.K., Greve, D.N., Salat, D.H., & Fischl, B. 2005. Regional cortical thinning in preclinical Huntington disease and its relationship to cognition. *Neurology*, 65, (5) 745-747
- Rosas, H.D., Koroshetz, W.J., Chen, Y.I., Skeuse, C., Vangel, M., Cudkovic, M.E., Caplan, K., Marek, K., Seidman, L.J., Makris, N., Jenkins, B.G., & Goldstein, J.M. 2003. Evidence for more widespread cerebral pathology in early HD: an MRI-based morphometric analysis. *Neurology*, 60, (10) 1615-1620
- Rosas, H.D., Liu, A.K., Hersch, S., Glessner, M., Ferrante, R.J., Salat, D.H., van der Kouwe, A., Jenkins, B.G., Dale, A.M., & Fischl, B. 2002. Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology*, 58, (5) 695-701
- Rosas, H.D., Reuter, M., Doros, G., Lee, S.Y., Triggs, T., Malarick, K., Fischl, B., Salat, D.H., & Hersch, S.M. 2011. A tale of two factors: what determines the rate of progression in Huntington's disease? A longitudinal MRI study. *Movement Disorders*, 26, (9) 1691-1697

- Rosas, H.D., Salat, D.H., Lee, S.Y., Zaleta, A.K., Pappu, V., Fischl, B., Greve, D., Hevelone, N., & Hersch, S.M. 2008. Cerebral cortex and the clinical expression of Huntington's disease: complexity and heterogeneity. *Brain*, 131, 1057-1068
- Rosas, H.D., Tuch, D.S., Hevelone, N.D., Zaleta, A.K., Vangel, M., Hersch, S.M., & Salat, D.H. 2006. Diffusion tensor imaging in presymptomatic and early Huntington's disease: Selective white matter pathology and its relationship to clinical measures. *Movement Disorders*, 21, (9) 1317-1325
- Rosenblatt, A., Kumar, B.V., Mo, A., Welsh, C.S., Margolis, R.L., & Ross, C.A. 2011. Age, CAG repeat length, and clinical progression in Huntington's disease. *Movement Disorders*.
- Rosenblatt, A., Liang, K.Y., Zhou, H., Abbott, M.H., Gourley, L.M., Margolis, R.L., Brandt, J., & Ross, C.A. 2006. The association of CAG repeat length with clinical progression in Huntington disease. *Neurology*, 66, (7) 1016-1020
- Ross, C.A. & Tabrizi, S.J. 2011. Huntington's disease: from molecular pathogenesis to clinical treatment. *Lancet Neurology*, 10, (1) 83-98
- Rothman, K.J. 1990. No adjustments are needed for Multiple Comparisons. *Epidemiology*, 1, (1) 43-46
- Rub, U., Hoche, F., Brunt, E.R., Heinsen, H., Seidel, K., Del, T.D., Paulson, H.L., Bohl, J., von, G.C., Vonsattel, J.P., Korf, H.W., & den Dunnen, W.F. 2013. Degeneration of the Cerebellum in Huntington's Disease (HD): Possible Relevance for the Clinical Picture and Potential Gateway to Pathological Mechanisms of the Disease Process. *Brain Pathology*, 23, (2) 165-177
- Rulseh, A.M., Keller, J., Tintera, J., Kozisek, M., & Vymazal, J. 2013. Chasing shadows: What determines DTI metrics in gray matter regions? An in vitro and in vivo study. *Journal of Magnetic Resonance Imaging*
- Ruocco, H.H., Bonilha, L., Li, L.M., Lopes-Cendes, I., & Cendes, F. 2008. Longitudinal analysis of regional grey matter loss in Huntington disease: effects of the length of the expanded CAG repeat. *Journal of Neurology, Neurosurgery and Psychiatry*, 79, (2) 130-135
- Ruocco, H.H., Lopes-Cendes, I., Li, L.M., Santos-Silva, M., & Cendes, F. 2006. Striatal and extrastriatal atrophy in Huntington's disease and its relationship with length of the CAG repeat. *Brazilian Journal of Medical & Biological Research*, 39, (8) 1129-1136
- Sakazume, S., Yoshinari, S., Oguma, E., Utsuno, E., Ishii, T., Narumi, Y., Shiihara, T., & Ohashi, H. 2009. A patient with early onset Huntington disease and severe cerebellar atrophy. *American Journal of Medical Genetics Part A*, 149A, (4) 598-601
- Sauter, D.A., Eisner, F., Calder, A.J., & Scott, S.K. 2010a. Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, 63, (11) 2251-2272
- Sauter, D.A., Eisner, F., Ekman, P., & Scott, S.K. 2010b. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Science U.S.A*, 107, (6) 2408-2412
- Scahill, R.I., Frost, C., Jenkins, R., Whitwell, J.L., Rossor, M.N., & Fox, N.C. 2003. A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. *Archives of Neurology*, 60, (7) 989-994
- Scahill, R.I., Hobbs, N.Z., Say, M.J., Bechtel, N., Henley, S.M., Hyare, H., Langbehn, D.R., Jones, R., Leavitt, B.R., Roos, R.A., Durr, A., Johnson, H., Lehericy, S., Craufurd, D., Kennard, C., Hicks, S.L., Stout, J.C.,

- Reilmann, R., & Tabrizi, S.J. 2011. Clinical impairment in premanifest and early Huntington's disease is associated with regionally specific atrophy. *Human Brain Mapping*.
- Scharmuller, W., Ille, R., & Schienle, A. 2013. Cerebellar Contribution to Anger Recognition Deficits in Huntington's Disease. *Cerebellum*
- Scherer, K.R. & Scherer, U. 2011. Assessing the Ability to Recognize Facial and Vocal Expressions of Emotion: Construction and Validation of the Emotion Recognition Index. *Journal of Nonverbal Behavior*, 35, (4) 305-326
- Schlegel, K., Grandjean, D., & Scherer, K.R. 2012. Emotion recognition: Unidimensional ability or a set of modality- and emotion-specific skills? *Personality and Individual Differences*, 53, (1) 16-21
- Schmahmann, J.D. & Caplan, D. 2006. Cognition, emotion and the cerebellum. *Brain*, 129, 290-292
- Schmahmann, J.D. & Sherman, J.C. 1998. The cerebellar cognitive affective syndrome. *Brain*, 121, 561-579
- Schott, J.M., Price, S.L., Frost, C., Whitwell, J.L., Rossor, M.N., & Fox, N.C. 2005. Measuring atrophy in Alzheimer disease: a serial MRI study over 6 and 12 months. *Neurology*, 65, (1) 119-124
- Segonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., & Fischl, B. 2004. A hybrid approach to the skull stripping problem in MRI. *Neuroimage*, 22, (3) 1060-1075
- Shepard, R. & Metzler, J. 1971. Mental rotation of three dimensional objects. *Science*, 171, (972) 701-703
- Shoulson, I. & Fahn, S. 1979. Huntington Disease - Clinical Care and Evaluation. *Neurology*, 29, (1) 1-3
- Simon, D., Craig, K.D., Gosselin, F., Belin, P., & Rainville, P. 2008. Recognition and discrimination of prototypical dynamic expressions of pain and emotions. *Pain*, 135, (1-2) 55-64
- Sled, J.G., Zijdenbos, A.P., & Evans, A.C. 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17, (1) 87-97
- Smith, A. 1991. *Symbol Digit Modalities Test* Los Angeles, Western Psychological Services.
- Smith, S.M. 2002. Fast robust automated brain extraction. *Human Brain Mapping*, 17, (3) 143-155
- Smith, S.M., De, S.N., Jenkinson, M., & Matthews, P.M. 2001. Normalized accurate measurement of longitudinal brain change. *Journal of Computer Assisted Tomography*, 25, (3) 466-475
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y.Y., De Stefano, N., Brady, J.M., & Matthews, P.M. 2004. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23, S208-S219
- Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., & De, S.N. 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage*, 17, (1) 479-489
- Snaith, R.P., Constantopoulos, A.A., Jardine, M.Y., & McGuffin, P. 1978. Clinical Scale for Self-Assessment of Irritability. *British Journal of Psychiatry*, 132, (FEB) 164-171
- Snowden, J.S., Austin, N.A., Sembi, S., Thompson, J.C., Craufurd, D., & Neary, D. 2008. Emotion recognition in Huntington's disease and frontotemporal dementia. *Neuropsychologia*, 46, (11) 2638-2649

- Song, S.K., Sun, S.W., Ju, W.K., Lin, S.J., Cross, A.H., & Neufeld, A.H. 2003. Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia. *Neuroimage*, 20, (3) 1714-1722
- Song, S.K., Sun, S.W., Ramsbottom, M.J., Chang, C., Russell, J., & Cross, A.H. 2002. Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. *Neuroimage*, 17, (3) 1429-1436
- Spreen, O. & Strauss, E. 1985. *A compendium of neuropsychological tests: Administration, norms and commentary*, 2nd ed. New York, Oxford University Press.
- Sprengelmeyer, R., Rausch, M., Eysel, U.T., & Przuntek, H. 1998. Neural structures associated with recognition of facial expressions of basic emotions. *Proceedings of Biological Science*, 265, (1409) 1927-1931
- Sprengelmeyer, R., Schroeder, U., Young, A.W., & Epplen, J.T. 2006. Disgust in pre-clinical Huntington's disease: a longitudinal study. *Neuropsychologia*, 44, (4) 518-533
- Sprengelmeyer, R., Young, A.W., Calder, A.J., Karnat, A., Lange, H., Homberg, V., Perrett, D.I., & Rowland, D. 1996. Loss of disgust. Perception of faces and emotions in Huntington's disease. *Brain*, 119 (Pt 5), 1647-1665
- Squitieri, F., Cannella, M., Simonelli, M., Sassone, J., Martino, T., Venditti, E., Ciammola, A., Colonnese, C., Frati, L., & Ciarmiello, A. 2009a. Distinct brain volume changes correlating with clinical stage, disease progression rate, mutation size, and age at onset prediction as early biomarkers of brain atrophy in Huntington's disease. *CNS Neuroscience & Therapeutics*, 15, (1) 1-11
- Squitieri, F., Orobello, S., Cannella, M., Martino, T., Romanelli, P., Giovacchini, G., Frati, L., Mansi, L., & Ciarmiello, A. 2009b. Riluzole protects Huntington disease patients from brain glucose hypometabolism and grey matter volume loss and increases production of neurotrophins. *European Journal of Nuclear Medicine & Molecular Imaging*, 36, (7) 1113-1120
- Sritharan, A., Egan, G.F., Johnston, L., Horne, M., Bradshaw, J.L., Bohanna, I., Asadi, H., Cunningham, R., Churchyard, A.J., Chua, P., Farrow, M., & Georgiou-Karistianis, N. 2010. A longitudinal diffusion tensor imaging study in symptomatic Huntington's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, 81, (3) 257-262
- Starkstein, S.E., Robinson, R.G., Berthier, M.L., & Price, T.R. 1988. Depressive-Disorders Following Posterior Circulation As Compared with Middle Cerebral-Artery Infarcts. *Brain*, 111, 375-387
- Stoffers, D., Sheldon, S., Kuperman, J.M., Goldstein, J., Corey-Bloom, J., & Aron, A.R. 2010. Contrasting gray and white matter changes in preclinical Huntington disease: an MRI study. *Neurology*, 74, (15) 1208-1216
- Sturrock, A., Laule, C., Decolongon, J., Santos, R.D., Coleman, A.J., Creighton, S., Bechtel, N., Reilmann, R., Hayden, M.R., Tabrizi, S.J., Mackay, A.L., & Leavitt, B.R. 2010. Magnetic resonance spectroscopy biomarkers in premanifest and early Huntington disease. *Neurology*, 75, (19) 1702-1710
- Sullivan, E.V. & Pfefferbaum, A. 2007. Neuroradiological characterization of normal adult ageing. *British Journal of Radiology*, 80 Spec No 2, S99-108
- Suzuki, A., Hoshino, T., & Shigemasa, K. 2014. Happiness is unique: A latent structure of emotion recognition traits revealed by statistical model comparison. *Personality and Individual Differences*, 48, 196-201

- Tabrizi, S.J., Langbehn, D.R., Leavitt, B.R., Roos, R.A., Durr, A., Craufurd, D., Kennard, C., Hicks, S.L., Fox, N.C., Scahill, R.I., Borowsky, B., Tobin, A.J., Rosas, H.D., Johnson, H., Reilmann, R., Landwehrmeyer, B., & Stout, J.C. 2009. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *Lancet Neurology*, 8, (9) 791-801
- Tabrizi, S.J., Reilmann, R., Roos, R.A., Durr, A., Leavitt, B., Owen, G., Jones, R., Johnson, H., Craufurd, D., Hicks, S.L., Kennard, C., Landwehrmeyer, B., Stout, J.C., Borowsky, B., Scahill, R.I., Frost, C., & Langbehn, D.R. 2012. Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data. *Lancet Neurology*
- Tabrizi, S.J., Scahill, R.I., Durr, A., Roos, R.A., Leavitt, B.R., Jones, R., Landwehrmeyer, G.B., Fox, N.C., Johnson, H., Hicks, S.L., Kennard, C., Craufurd, D., Frost, C., Langbehn, D.R., Reilmann, R., & Stout, J.C. 2011. Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: the 12-month longitudinal analysis. *Lancet Neurology*, 10, (1) 31-42
- Tabrizi, S.J., Scahill, R.I., Owen, G., Durr, A., Leavitt, B.R., Roos, R.A., Borowsky, B., Landwehrmeyer, B., Frost, C., Johnson, H., Craufurd, D., Reilmann, R., Stout, J.C., & Langbehn, D.R. 2013. Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. *Lancet Neurology*, 12, (7) 637-649
- Talairach, J., Tournoux, P., & Musolino, A. 1988. Anatomical Stereotaxic Studies of the Frontal-Lobe in the Management of the Epilepsies. *Epilepsia*, 29, (2) 205
- Tassicker, R.J., Teltscher, B., Trembath, M.K., Collins, V., Sheffield, L.J., Chiu, E., Gurrin, L., & Delatycki, M.B. 2009. Problems assessing uptake of Huntington disease predictive testing and a proposed solution. *European Journal of Human Genetics*, 17, (1) 66-70
- Thacker, N.A., Jackson, A., Moriarty, D., & Vokurka, E. 1999. Improved quality of re-sliced MR images using re-normalized sinc interpolation. *Jmri-Journal of Magnetic Resonance Imaging*, 10, (4) 582-588
- Tootell, R.B., Hadjikhani, N.K., Vanduffel, W., Liu, A.K., Mendola, J.D., Sereno, M.I., & Dale, A.M. 1998. Functional analysis of primary visual cortex (V1) in humans. *Proceedings of the National Academy of Science U.S.A.*, 95, (3) 811-817
- Tootell, R.B., Tsao, D., & Vanduffel, W. 2003. Neuroimaging weighs in: humans meet macaques in "primate" visual cortex. *Journal of Neuroscience*, 23, (10) 3981-3989
- Vandenberghe, W., Demaerel, P., Dom, R., & Maes, F. 2009. Diffusion-weighted versus volumetric imaging of the striatum in early symptomatic Huntington disease. *Journal of Neurology*, 256, (1) 109-114
- Vonsattel, J.P., Myers, R.H., Stevens, T.J., Ferrante, R.J., Bird, E.D., & Richardson, E.P. 1985. Neuropathological Classification of Huntingtons-Disease. *Journal of Neuropathology and Experimental Neurology*, 44, (6) 559-577
- Vonsattel, J.P.G. & DiFiglia, M. 1998. Huntington disease. *Journal of Neuropathology and Experimental Neurology*, 57, (5) 369-384
- Wakana, S., Caprihan, A., Panzenboeck, M.M., Fallon, J.H., Perry, M., Gollub, R.L., Hua, K.G., Zhang, J.Y., Jiang, H.Y., Dubey, P., Blitz, A., van Zijl, P., & Mori, S. 2007. Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage*, 36, (3) 630-644

- Walker, A., Liney, G., Metcalfe, P., & Holloway, L. 2014. MRI distortion: considerations for MRI based radiotherapy treatment planning. *Australasian Physical and Engineering Science in Medicine*, 37, (1) 103-113
- Wandell, B.A., Dumoulin, S.O., & Brewer, A.A. 2007. Visual field maps in human cortex. *Neuron*, 56, (2) 366-383
- Wang, F., Jackowski, M., Kalmar, J.H., Chepenik, L.G., Tie, K., Qiu, M., Gong, G.L., Pittman, B.P., Jones, M.M., Shah, M.P., Spencer, L., Papademetris, X., Constable, R.T., & Blumberg, H.P. 2008. Abnormal anterior cingulum integrity in bipolar disorder determined through diffusion tensor imaging. *British Journal of Psychiatry*, 193, (2) 126-129
- Wang, K., Hoosain, R., Yang, R.M., Meng, Y., & Wang, C.Q. 2003. Impairment of recognition of disgust in Chinese with Huntington's or Wilson's disease. *Neuropsychologia*, 41, (5) 527-537
- Warfield, S.K., Zou, K.H., & Wells, W.M. 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23, (7) 903-921
- Warner, J. H. & Hayden, M. A new model for age-at-onset in Huntington's Disease, Presented in *CHDI Foundation Huntington's Disease Therapeutics Conference (Palm Springs, CA, USA)*.
- Weaver, K.E., Richards, T.L., Liang, O., Laurino, M.Y., Samii, A., & Aylward, E.H. 2009. Longitudinal diffusion tensor imaging in Huntington's Disease. *Experimental Neurology*, 216, (2) 525-529
- Weir, D.W., Sturrock, A., & Leavitt, B.R. 2011. Development of biomarkers for Huntington's disease. *Lancet Neurology*, 10, (6) 573-590
- Wexler, N.S., Lorimer, J., Porter, J., Gomez, F., Moskowitz, C., Shackell, E., Marder, K., Penchaszadeh, G., Roberts, S.A., Gayan, J., Brocklebank, D., Cherny, S.S., Cardon, L.R., Gray, J., Dlouhy, S.R., Wiktorski, S., Hodes, M.E., Conneally, P.M., Penney, J.B., Gusella, J., Cha, J.H., Irizarry, M., Rosas, D., Hersch, S., Hollingsworth, Z., MacDonald, M., Young, A.B., Andresen, J.M., Housman, D.E., De Young, M.M., Bonilla, E., Stillings, T., Negrette, A., Snodgrass, S.R., Martinez-Jaurrieta, M.D., Ramos-Arroyo, M.A., Bickham, J., Ramos, J.S., Marshall, F., Shoulson, I., Rey, G.J., Feigin, A., Arnheim, N., Acevedo-Cruz, A., Acosta, L., Alvir, J., Fischbeck, K., Thompson, L.M., Young, A., Dure, L., O'Brien, C.J., Paulsen, J., Brickman, A., Krch, D., Peery, S., Hogarth, P., Higgins, D.S., Jr., & Landwehrmeyer, B. 2004. Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proceedings of the National Academy of Science U.S.A*, 101, (10) 3498-3503
- Whittaker, J.F., Deakin, J.F.W., & Tomenson, B. 2001. Face processing in schizophrenia: defining the deficit. *Psychological Medicine*, 31, (3) 499-507
- Whitwell, J.L., Crum, W.R., Watt, H.C., & Fox, N.C. 2001. Normalization of cerebral volumes by use of intracranial volume: Implications for longitudinal quantitative MR imaging. *American Journal of Neuroradiology*, 22, (8) 1483-1489
- Wild, E.J., Henley, S.M., Hobbs, N.Z., Frost, C., MacManus, D.G., Barker, R.A., Fox, N.C., & Tabrizi, S.J. 2010. Rate and acceleration of whole-brain atrophy in premanifest and early Huntington's disease. *Movement Disorders*, 25, (7) 888-895
- Wise, J. 2010. Prevalence of Huntington's disease is underestimated in UK. *British Medical Journal*, 340,

- Wolf, R.C., Sambataro, F., Vasic, N., Baldas, E.M., Ratheiser, I., Bernhard, L.G., Depping, M.S., Thomann, P.A., Sprengelmeyer, R., Sussmuth, S.D., & Orth, M. 2014. Visual system integrity and cognition in early Huntington's disease. *European Journal of Neuroscience*
- Wolf, R.C., Thomann, P.A., Thomann, A.K., Vasic, N., Wolf, N.D., Landwehrmeyer, G.B., & Orth, M. 2013. Brain Structure in Preclinical Huntington's Disease: A Multi-Method Approach. *Neurodegenerative Diseases*, 12, (1) 13-22
- Wolf, R.C., Vasic, N., Schonfeldt-Lecuona, C., Landwehrmeyer, G.B., & Ecker, D. 2007. Dorsolateral prefrontal cortex dysfunction in presymptomatic Huntington's disease: evidence from event-related fMRI. *Brain*, 130, (Pt 11) 2845-2857
- Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., & Mazziotta, J.C. 1998a. Automated image registration: I. General methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography*, 22, (1) 139-152
- Woods, R.P., Grafton, S.T., Watson, J.D.G., Sicotte, N.L., & Mazziotta, J.C. 1998b. Automated image registration: II. Intersubject validation of linear and nonlinear models. *Journal of Computer Assisted Tomography*, 22, (1) 153-165
- Woodworth, R.S. & Schlosberg, H. 1954. *Experimental psychology: revised edition* New York, Henry Holt.
- Wu, T. & Hallett, M. 2013. The cerebellum in Parkinson's disease. *Brain*, 136, 696-709
- Zeki, S.M. 1978. Functional specialisation in the visual cortex of the rhesus monkey. *Nature*, 274, (5670) 423-428
- Zhang, H., Avants, B.B., Yushkevich, P.A., Woo, J.H., Wang, S., McCluskey, L.F., Elman, L.B., Melhem, E.R., & Gee, J.C. 2007. High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: an example study using amyotrophic lateral sclerosis. *IEEE Transactions in Medical Imaging*, 26, (11) 1585-1597
- Zhang, Y., Brady, M., & Smith, S. 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions in Medical Imaging*, 20, (1) 45-57
- Zhang, Y. & Friedlander, R.M. 2011. Using non-coding small RNAs to develop therapies for Huntington's disease. *Gene Therapy*, 18, (12) 1139-1149
- Zhang, Y., Long, J.D., Mills, J.A., Warner, J.H., Lu, W.J., & Paulsen, J.S. 2011. Indexing Disease Progression at Study Entry With Individuals At-Risk for Huntington Disease. *American Journal of Medical Genetics Part B- Neuropsychiatric Genetics*, 156B, (7) 751-763
- Zigmond, A.S. & Snaith, R.P. 1983. The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, 67, (6) 361-370